

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

NADIA CAROLINA DE AGUIAR FRACASSO

**DIVERSIDADE GENÉTICA DAS REGIÕES REGULATÓRIAS E CODIFICANTES DOS
GENES *SLC45A2* E *TYR* EM AMOSTRA DA POPULAÇÃO BRASILEIRA**

**Ribeirão Preto
2018**

NADIA CAROLINA DE AGUIAR FRACASSO

**DIVERSIDADE GENÉTICA DAS REGIÕES REGULATÓRIAS E CODIFICANTES DOS
GENES *SLC45A2* E *TYR* EM AMOSTRA DA POPULAÇÃO BRASILEIRA**

Tese apresentada ao Programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto para a obtenção do título de Doutora em Ciências.

Área de concentração:
Genética

Orientador:
Prof. Dr. Celso Teixeira Mendes Junior

Ribeirão Preto
2018

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.

FICHA CATALOGRÁFICA

Fracasso, Nadia Carolina de Aguiar.

Diversidade genética das regiões regulatórias e codificantes dos genes *SLC45A2* e *TYR* em amostra da população brasileira / Nádia Carolina de Aguiar Fracasso. Ribeirão Preto, 2018.

188 f.: il. 9; 30 cm.

Tese (Doutorado em Ciências) – Faculdade de Medicina de Ribeirão Preto / Universidade de São Paulo. Área de concentração: Genética.

Orientador: Mendes-Junior, Celso Teixeira.

1. *SLC45A2*, 2. *TYR*, 3. SNPs, 4. Pigmentação, 5. População brasileira

FICHA DE APROVAÇÃO

Nome: FRACASSO, Nadia Carolina de Aguiar

Título: Diversidade genética das regiões regulatórias e codificantes dos genes *SLC45A2* e *TYR* em amostra da população brasileira

Tese apresentada ao Programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto para a obtenção do título de Doutora em Ciências, área de concentração Genética.

Aprovado em: ___ / ___ / ___

Banca Examinadora:

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Dedico esta tese aos meus pais José Francisco e Enide (*in memoriam*), a toda minha família maravilhosa e aos meus amigos. Obrigada pelo amor, apoio e por sempre acreditarem em mim.

AGRADECIMENTOS

Ao meu pai, pelo apoio moral e financeiro, pelo amor e paciência. Obrigada por estar sempre ao meu lado.

A minha mãe (*in memoriam*), que me tornou um ser humano melhor com seus bons exemplos e que me faz refletir todos os dias sobre as coisas mais importantes na vida. Obrigada por sempre olhar por mim, sei que nos encontraremos de novo.

A minha irmã, por compartilhar sua sabedoria, pelo incentivo, por sempre acreditar em mim e por ter me presenteado com sobrinhos tão maravilhosos.

Ao meu irmão, por se tornar um exemplo para mim de que as pessoas sempre podem ser melhores, pela família linda que construiu, pelo incentivo e apoio nas minhas conquistas. Você é meu orgulho!

Ao Prof. Dr. Celso Teixeira Mendes Junior, pela orientação neste trabalho, por acreditar na minha capacidade, pelos inúmeros incentivos e estímulos, pela compreensão, paciência, atenção e boa vontade. Obrigada pela jornada e por ser um ser humano tão incrível.

Aos meus familiares pelo amor, carinho, estímulo e pelas infinitas horas de alegria que passamos juntos.

Aos amigos do Laboratório de Pesquisas Forenses e Genômicas: Guilherme D., Maria Luiza, Alison, Guilherme IC, Letícia, Thássia, Amanda e ao técnico do laboratório, André. Obrigada pela jornada juntos!! Pelo conhecimento compartilhado, pelos momentos de descontração e parceria e, principalmente, por me motivarem a melhorar sempre. Vocês são o melhor Lab!! Em especial a Malu e Alison que me ajudaram demais ao longo do caminho e nos momentos finais. Obrigada!!

As queridas amigas Maria Luiza (Malu) e Marcela (Marida), por compartilharem o conhecimento, pelos ensinamentos, conselhos, paciência, conversas, diversão, e, principalmente, pela ajuda emocional no desenvolvimento dessa tese. Eu não teria conseguido sem vocês. Gratidão!!

As amigas de faculdade (Veronica, Monica e Ludmila) e a amiga Vanessa Borges (você é luz) por estarem sempre presentes mesmo longe, por dividirem anos e momentos incríveis comigo, pelo apoio e carinho sempre.

A minha querida amiga, nutricionista, acupunturista e parceira de vivências Eliane por todo apoio emocional, espiritual e energético. Você é luz!! Gratidão!!

Ao Prof. Dr. Aguinaldo Luiz Simões por gentilmente ceder à utilização dos equipamentos de seu laboratório. E aos técnicos Cláudia, Elisabete e Rubens pelo auxílio prestado na realização deste trabalho.

Ao Laboratório de Imunopatologia Molecular, em especial ao Prof. Dr. Eduardo Antônio Donadi, pelo suporte e estrutura para realização deste trabalho e à Juliana Doblás Massaro, Sandra e Flávia, que auxiliaram na realização dos ensaios de sequenciamento.

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

À Profa. Milena e às enfermeiras Thalita, Ana Paula, Lorena e Marina por nos ajudarem nas coletas.

Aos professores membros da banca examinadora, pela disponibilidade para avaliar este trabalho.

Aos professores e funcionários do Departamento de Genética, pelos ensinamentos, paciência e apoio. Em especial à secretária, Susie Nalon, por toda gentileza no atendimento e disponibilidade sempre.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro recebido para o desenvolvimento do trabalho.

A todos os doadores voluntários, sem os quais essa pesquisa não poderia ser desenvolvida.

A todas as pessoas que de alguma forma contribuíram para a realização deste trabalho,
MUITO OBRIGADA!

*“Love, love, love
Love, love, love
Love, love, love*

*There's nothing you can do that can't be done
Nothing you can sing that can't be sung
Nothing you can say but you can learn how to play the game
It's easy*

*There's nothing you can make that can't be made
No one you can save that can't be saved
Nothing you can do but you can learn how to be you in time
It's easy*

*All you need is love
All you need is love
All you need is love, love
Love is all you need*

*Love, love, love
Love, love, love
Love, love, love*

*All you need is love
All you need is love
All you need is love, love
Love is all you need*

*There's nothing you can know that isn't known
Nothing you can see that isn't shown
No where you can be that isn't where you're meant to be
It's easy*

*All you need is love
All you need is love
All you need is love, love
Love is all you need*

*All you need is love
All you need is love
All you need is love, love
Love is all you need”*

*All you need is love
The Beatles*

RESUMO

FRACASSO, N. C. A. **Diversidade genética das regiões regulatórias e codificantes dos genes *SLC45A2* e *TYR* em amostra da população brasileira.** 2018. 188 páginas. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brasil.

O gene *SLC45A2* codifica a proteína MATP, envolvida na síntese de melanina através do processamento e transporte intracelular da tirosinase e transporte de prótons para o melanosomo. Por sua vez, a tirosinase, codificada pelo gene *TYR*, catalisa os dois primeiros passos da conversão de tirosina em melanina, além de atuar em uma etapa final da biossíntese de eumelanina. Considerando que polimorfismos nestes genes influenciam a variação de características normais de pigmentação, que somente recentemente mais informações sobre a função do gene *SLC45A2* foram descobertas e que diversas questões sobre o envolvimento da tirosinase nos fenótipos normais de pigmentação ainda permanecem sem total compreensão, é necessário o melhor entendimento das interações entre as variantes desses genes de pigmentação e moléculas regulatórias, assim como o conhecimento sobre a diversidade de tais genes em uma população miscigenada como a brasileira. Dessa forma, o presente trabalho teve como objetivo analisar a diversidade genética das regiões, promotora, codificante e 3'UTR dos genes *SLC45A2* e *TYR* em uma amostra da população brasileira. As regiões regulatórias e codificadoras dos genes *SLC45A2* e *TYR* foram analisadas por sequenciamento de nova geração em uma amostra de 340 indivíduos, estratificados de acordo com a pigmentação dos olhos, cabelos e pele, bem como quanto à presença ou ausência de sardas. Bibliotecas de DNA foram preparadas utilizando o *Haloplex Target Enrichment System (Agilent Technologies)* e sequenciadas por meio da plataforma *MiSeq (Illumina)*. Os softwares cutadapt, BWA e GATK/VCFx foram utilizados para trimagem dos adaptadores, alinhamento e chamada de variantes, respectivamente. O programa PHASE foi utilizado para a reconstrução dos haplótipos. Um total de 58 variantes foram identificadas no gene *SLC45A2*. Destas, 28 foram associadas a pelo menos uma das características de pigmentação avaliadas. Embora as regiões não-codificadoras tenham apresentado maior diversidade genética, associações significativas também foram encontradas em regiões exônicas. Dentre estas, destaca-se a associação do alelo rs16891982*G (exon 5, Leu374Phe) com pele clara ($p = 9,54 \times 10^{-35}$, OR = 22,66) e cabelos loiros ($p = 1,59 \times 10^{-26}$, OR = 31,72). Quando observamos os haplótipos inferidos para as regiões codificantes das isoformas 1 e 2 que possuem esse SNP e as associações encontradas, podemos notar que os únicos haplótipos associados com fenótipos de pigmentação claros para ambas isoformas ["Iso2cds1": pele clara ($p = 3,65 \times 10^{-29}$, OR =

22,63) e cabelos loiros ($p = 6,34 \times 10^{-23}$, OR = 26,19); “Iso1cds1”: pele clara ($p = 1,08 \times 10^{-24}$, OR = 17,38) e cabelos loiros ($p = 1,94 \times 10^{-22}$, OR = 22,93)] possuem como diferença em relação a todos os outros haplótipos associados com pigmentação escura o alelo rs16891982*G. Para o gene *TYR* foram identificadas 42 variantes e 15 se mostraram associadas a algum dos fenótipos de pigmentação avaliados. A maior parte da diversidade desse gene foi encontrada na região intrônica, com ênfase para a associação do genótipo rs1393350*A/A com olhos azuis ($p = 0,0253$, OR = 13,06) e cabelos castanho-claros ($p = 0,0019$, OR = 16,07). Ao analisarmos as associações encontradas para os haplótipos inferidos para essa região, podemos notar que os haplótipos “cds5” ($p = 1,71 \times 10^{-05}$, OR = 21,26), “cds7” ($p = 0,0061$, OR = 23,70) e “cds9” ($p = 0,0017$, OR = 29,25) estão associados a pele escura e o haplótipo “cds12” ($p = 0,015$, OR = 21,61) a ausência de sardas. Quando nos atentamos a composição desses haplótipos em relação ao SNP rs1393350, podemos perceber que todos os haplótipos possuem o alelo referência (rs1393350*G), o que é consistente com a associação entre o genótipo (rs1393350*A/A) e fenótipos de pigmentação claros. Os resultados aqui encontrados reafirmam a importância desempenhada pelos genes *SLC45A2* e *TYR* na geração da diversidade de fenótipos de pigmentação.

Palavras-chave: SLC45A2, TYR, SNPs, Pigmentação, População brasileira.

ABSTRACT

FRACASSO, N. C. F. **Genetic diversity of the regulatory and coding regions of *SLC45A2* and *TYR* genes in a Brazilian population sample.** 2018. 188 pages. Thesis (PhD Degree) – Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil.

The *SLC45A2* gene encodes the Membrane-Associated Transporter Protein, which mediates melanin synthesis by tyrosinase trafficking and proton transportation to melanosomes. On the other hand, the tyrosinase protein, encoded by the *TYR* gene, catalyzes the first two steps of tyrosine to melanin conversion in addition to acting in a final stage of eumelanin biosynthesis. Considering that polymorphisms in these genes influence normal pigmentation variation, that only recently more information about *SLC45A2* gene function were discovered and that many questions about the tyrosinase involvement in normal pigmentation phenotypes are still not fully understood, it is necessary to better understand the interactions between variants in these pigmentation genes and regulatory molecules, as well as to improve knowledge about their diversity in a mixed population like the Brazilian one. Thus, the present study aimed at analyzing the genetic diversity of the promoter, coding and 3'UTR regions from the *SLC45A2* and *TYR* genes in a Brazilian admixed population sample ($n=340$). The regulatory and coding regions were analyzed by next-generation sequencing procedures. The individuals were stratified according to eye, hair and skin pigmentation, as well as to the presence or absence of freckles. DNA libraries were prepared using the *Haloplex* Target Enrichment System (Agilent Technologies) and sequenced at the MiSeq platform (Illumina). cutadapt, BWA and GATK/VCFx software packages were used for trimming adaptor sequences, alignment and genotype calling, respectively. The PHASE program was used for haplotypes reconstruction. A total of 58 variation sites were identified in the *SLC45A2* gene. Of these, 28 were found in association with at least one of the analyzed pigmentation characteristics. Although the non-coding regions were more diverse, the exonic region also showed significant associations. Among them, the association of the rs16891982*G allele (exon 5, Leu374Phe) with light skin ($p = 9.54 \times 10^{-35}$, OR = 22.66) and blond hair ($p = 1.59 \times 10^{-26}$, OR = 31.72) stands out. When we observe the inferred haplotypes for the isoforms 1 and 2 coding regions that have this SNP and the associations found, we can recognize that haplotypes associated with light pigmentation phenotypes ["Iso2cds1": light skin ($p = 3.65 \times 10^{-29}$, OR = 22.63) and blond hair ($p = 6.34 \times 10^{-23}$, OR = 26.19); "Iso1cds1": light skin ($p = 1.08 \times 10^{-24}$, OR = 17.38) and blond hair ($p = 1.94 \times 10^{-22}$, OR = 22.93)] have the rs16891982*G allele, while haplotypes associated with dark pigmentation harbors the other one. Forty-two variation sites were

identified for the *TYR* gene and 15 of them were associated with one of the evaluated pigmentation phenotypes. Most of the diversity of this gene was found in the intronic region, with emphasis on the association of genotype rs1393350*A/A with blue eyes ($p = 0.0253$, OR = 13.06) and light brown hair ($p = 0.0019$, OR = 16.07). When we analyze the associations found for the inferred haplotypes for this region, we can note that the haplotypes "cds5" ($p = 1.71 \times 10^{-05}$, OR = 21.25), "cds7" ($p = 0.0061$, OR = 23.70) and "cds9" ($p = 0.0017$, OR = 29.25) were associated with dark skin and the haplotype "cds12" ($p = 0.015$, OR = 21.61) with absence of freckles. When the composition of the haplotypes concerning this SNP (rs1393350) is taken into account, we can see that all haplotypes have the reference allele (rs1393350*G), which is consistent with the association between the rs1393350*A/A genotype and lighter pigmentation phenotypes. These findings provide additional support to the role played by *SLC45A2* and *TYR* in the generation of pigmentation phenotypes diversity.

Keywords: SLC45A2, TYR, SNPs, Pigmentation, Brazilian population.

LISTA DE ILUSTRAÇÕES

FIGURA 1. Esquema da estrutura da pele humana: tipos de pigmentação clara (*light*) e escura (*dark*). De cima para baixo: UV = radiação ultravioleta; SC = estrato córneo; G = estrato granuloso; S = estrato espinhoso; B = estrato basal; BM = membrana basal; D = derme. Tipos celulares: K = queratinócitos; M = melanócitos; F = fibroblastos; grânulos ovais = melanossomos contendo melanina (Yamaguchi, et al. 2007) _____ 08

FIGURA 2. Esquema da função transportadora da proteína MATP na regulação do pH melanossomal em situação normal e na ocorrência de OCA4. *Proton pump* = bomba de prótons; *Copper pump* = bomba de cobre; Cu = cobre; *Apo-Tyrosinase* = tirosinase sem cobre (atividade reduzida); *Cu-Tyrosinase* = tirosinase com cobre (ativada) (Bin, et al. 2015) _____ 14

FIGURA 3. Esquema da interação entre as proteínas de membrana (TYR, TYRP1 e OCA2) com adaptadores proteicos (AP) e posterior transporte para as membranas melanossomais através de vesículas isoladas (no caso da TYR) ou de túbulos que se originam nos endossomos e se fundem com a membrana de melanossomos próximos (adaptada de Sitaram, et al. 2012) _____ 18

FIGURA 4. Representação do gene TYR no programa IGV_2.3.66. A parte principal da tela de visualização é dividida em 2 partes. No topo pode ser vista a representação esquemática do cromossomo 11 e a região deste cromossomo que está sendo visualizada (indicada por uma barra vermelha). Logo abaixo, é apresentada uma escala contendo as coordenadas da região ampliada. Na parte inferior da tela temos a anotação do genoma para a região visualizada (*RefSeq Genes*) em azul e em verde, a anotação das regiões cobertas pelo nosso ensaio _____ 24

FIGURA 5. Representação do gene SLC45A2 no programa IGV_2.3.66. A parte principal da tela de visualização é dividida em 2 partes. No topo pode ser vista a representação esquemática do cromossomo 5 e a região deste cromossomo que está sendo visualizada (indicada por uma barra vermelha). Logo abaixo, é apresentada uma escala contendo as coordenadas da região ampliada. Na parte inferior da tela temos a anotação do genoma para a região visualizada (*RefSeq Genes*) em azul com as três isoformas do gene analisadas neste trabalho (acima a isoforma com sete éxons, no meio a isoforma com seis éxons e abaixo a isoforma com quatro éxons). Em verde, a anotação das regiões cobertas pelo nosso ensaio _____ 25

FIGURA 6. Representação dos alinhamentos das reads na região do gene SLC45A2 no programa IGV_2.3.66. A parte principal da tela de visualização é dividida em 3 partes. No topo pode ser vista a representação esquemática do cromossomo 5 e a região deste cromossomo que está sendo visualizada (indicada por uma barra vermelha). Logo abaixo, é apresentada uma escala contendo as coordenadas da região ampliada. Na parte intermediária da tela é mostrada parte do arquivo BAM carregado. A visualização desse arquivo é dividida em duas regiões; acima, temos um gráfico representativo da cobertura (ou número de *reads* cobrindo cada base do genoma) e abaixo, os alinhamentos das *reads*. Na parte inferior da tela temos a anotação do genoma para a região visualizada (*RefSeq Genes*) em azul e em verde, a anotação das regiões cobertas pelo nosso ensaio _____ 30

FIGURA 7. Representação esquemática das isoformas do gene SLC45A2. Os retângulos azuis representam os éxons e as regiões promotora (à esquerda) e 3'UTR (à direita). A isoforma 1 possui a 3'UTR de menor tamanho (31pb). A isoforma 2 possui 15pb a mais no éxon 6 e a 3'UTR de maior tamanho (1280pb). A isoforma 3

não possui os éxons 3, 6 e 7; o éxon 5 tem apenas 25pb e sua 3'UTR tem 245pb (incluindo parte do éxon 5 das outras duas isoformas). A figura não está em escala

32

FIGURA 8. Representação do desequilíbrio de ligação para os pontos de variação do gene *SLC45A2*, considerando a região promotora no lado direito da imagem. O bloco em desequilíbrio está delimitado em preto. Vermelho indica LD completo forte [$LOD (Log\ of\ the\ Odds) \geq 2$ e $D' = 1$]; azul indica LD completo fraco [$LOD < 2$ e $D' = 1$]; tons de rosa indicam LD moderado [$LOD \geq 2$ e $D' < 1$]; branco indica ausência de LD [$LOD < 2$ e $D' < 1$]. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). Valores de D' diferentes de 1 estão representados na forma de porcentagem no interior dos quadros com tons de rosa ou brancos

52

FIGURA 9. Representação do desequilíbrio de ligação para os pontos de variação do gene *TYR*. O bloco em desequilíbrio está delimitado em preto. Vermelho indica LD completo forte [$LOD (Log\ of\ the\ Odds) \geq 2$ e $D' = 1$]; azul indica LD completo fraco [$LOD < 2$ e $D' = 1$]; tons de rosa indicam LD moderado [$LOD \geq 2$ e $D' < 1$]; branco indica ausência de LD [$LOD < 2$ e $D' < 1$]. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). Valores de D' diferentes de 1 estão representados na forma de porcentagem no interior dos quadros com tons de rosa ou brancos

53

LISTA DE TABELAS

- Tabela 1:** Características da amostra populacional de acordo com a heteroclassificação_____36
- Tabela 2:** Dados do aproveitamento e qualidade dos cinco ensaios de sequenciamento_____37
- Tabela 3:** Haplótipos encontrados para região promotora do gene *SLC45A2* considerando 24 pontos de variação em 331 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho_____ 38
- Tabela 4:** Haplótipos encontrados para região codificante da isoforma 1 do gene *SLC45A2* considerando 32 pontos de variação em 307 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho_____39
- Tabela 5:** Haplótipos encontrados para região codificante da isoforma 2 do gene *SLC45A2* considerando 20 pontos de variação em 307 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho_____41
- Tabela 6:** Haplótipos encontrados para região codificante da isoforma 3 do gene *SLC45A2* considerando 18 pontos de variação em 310 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho_____42
- Tabela 7:** Haplótipos encontrados para região 3'UTR da isoforma 2 do gene *SLC45A2* considerando 9 pontos de variação em 337 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho_____ 43
- Tabela 8:** Haplótipos encontrados para região 3'UTR da isoforma 1 do gene *SLC45A2* considerando dois pontos de variação em 340 amostras_____ 43
- Tabela 9:** Haplótipos encontrados para região promotora do gene *TYR* considerando 20 pontos de variação em 305 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho_____44
- Tabela 10:** Haplótipos encontrados para região codificante do gene *TYR* considerando 21 pontos de variação em 327 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho_____45
- Tabela 11:** Frequências alélicas, heterozigose observada (H_o), heterozigose esperada (H_e) e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) para cada *locus* do gene *SLC45A2*. Valores significativos ($\alpha = 0,05$) na análise de aderência ao equilíbrio estão marcados em vermelho_____47
- Tabela 12:** Frequências alélicas, heterozigose observada (H_o), heterozigose esperada (H_e) e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) para cada *locus* do gene *TYR*. Valores significativos ($\alpha = 0,05$) na análise de aderência ao equilíbrio estão marcados em vermelho_____49
- Tabela 13:** Alelos e genótipos de SNPs/Indels do gene *SLC45A2* associados com as características fenotípicas avaliadas_____54

Tabela 14: Alelos e genótipos de SNPs/Indels do gene *TYR* associados com as características fenotípicas avaliadas_____58

Tabela 15: Haplótipos formados por SNPs/indels das regiões avaliadas para o gene *SLC45A2* associados com as características fenotípicas avaliadas_____60

Tabela 16: Haplótipos formados por SNPs/indels das regiões avaliadas para o gene *TYR* associados com as características fenotípicas avaliadas_____62

Tabela 17: Distribuição dos índices de melanina de acordo com os genótipos, considerando os 28 SNPs do gene *SLC45A2* previamente associados a alguma das características de pigmentação avaliadas (Tabela 14). R/R representam indivíduos com o genótipo homocigoto alelo referência, R/A são indivíduos com genótipo heterocigoto e A/A são indivíduos com genótipo homocigoto para o alelo alternativo_____63

Tabela 18: Distribuição dos índices de melanina de acordo com os genótipos, considerando os 15 SNPs do gene *TYR* previamente associados a alguma das características de pigmentação avaliadas (Tabela 13). R/R representam indivíduos com o genótipo homocigoto alelo referência, R/A são indivíduos com genótipo heterocigoto e A/A são indivíduos com genótipo homocigoto para o alelo alternativo_____64

Tabela 19: Relação dos 20 miRNAs que apresentaram maiores valores de ΔE , ΔS e ΔES para a região 3'UTR com apenas um SNP (rs16891982) do gene *SLC45A2*_____65

Tabela 20: Relação dos 20 miRNAs que apresentaram maiores valores de ΔE , ΔS e ΔES para os três haplótipos da região 3'UTR isoforma 1 do gene *SLC45A2*_____67

Tabela 21: Relação dos 20 miRNAs que apresentaram maiores valores de ΔE , ΔS e ΔES para os sete haplótipos da região 3'UTR isoforma 2 do gene *SLC45A2*_____68

Tabela 22: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados na região promotora do gene *SLC45A2*, nas populações mundiais e na população do presente estudo_____73

Tabela 23: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados nas regiões codificante e 3'UTR do gene *SLC45A2*, nas populações mundiais e na população do presente estudo_____75

Tabela 24: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados na região promotora do gene *TYR*, nas populações mundiais e na população do presente estudo_____81

Tabela 25: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados na região codificante e *downstream* do gene *TYR*, nas populações mundiais e na população do presente estudo_____84

SUMÁRIO

1. INTRODUÇÃO	7
1.1 Biologia e Genética da Pigmentação Humana	7
1.1.1 Impacto funcional de mutações em genes de pigmentação.....	11
1.2 <i>SLC45A2</i>	12
1.3 <i>TYR</i>	16
1.4 População Brasileira	20
2. HIPÓTESE	21
3. JUSTIFICATIVA	21
4. OBJETIVO.....	22
5. MATERIAL E MÉTODOS	23
5.1 Delineamento Experimental.....	23
5.2 Amostra Populacional	26
5.3 Determinação Fenotípica dos Indivíduos.....	26
5.4 Análise Laboratorial	27
5.4.1 Extração de DNA.....	27
5.4.2 Quantificação e verificação da qualidade do DNA genômico	28
5.4.3 Preparação das Bibliotecas.....	28
5.4.4 Sequenciamento MiSeq	28
5.5 Análises de Dados e Estatísticas.....	29
5.5.1 Pré-processamento e chamada de variantes.....	29
5.5.2 Inferência computacional de haplótipos	31
5.5.3 Análises estatísticas.....	33
5.5.4 Análise dos índices de melanina.....	33
5.5.5 Análise de miRNAs	34
6. RESULTADOS	35
6.1 Caracterização populacional e sequenciamento	35
6.2 Chamada de variantes e haplótipos inferidos	37
6.3 Análises populacionais e desequilíbrio de ligação	46
6.4 Associações fenotípicas	54
6.5 Associações índices de melanina.....	62
6.6 Análise de miRNAs	65
7. DISCUSSÃO	70
7.1 Aspectos populacionais e desvios do equilíbrio de Hardy-Weinberg.....	70
7.2 Associações encontradas.....	71
7.2.1 Gene <i>SLC45A2</i>	72

7.2.2 Gene <i>TYR</i>	81
7.3 Considerações finais.....	86
8. CONCLUSÕES	87
9. REFERÊNCIAS BIBLIOGRÁFICAS	88
10. ANEXOS	98
10.1 ANEXO A.....	98
10.2 ANEXO B.....	100
10.3 ANEXO C.....	102
10.4 ANEXO D.....	103
10.5 ANEXO E.....	111
10.6 ANEXO F.....	129
10.7 ANEXO G	139
11. MANUSCRITO	144

1. INTRODUÇÃO

A infinidade de diferentes tons de pele existentes hoje surgiu quando os humanos modernos saíram da África e se adaptaram a novos ambientes. Bem antes disso, os ancestrais do homem moderno possuíam pouca ou nenhuma variabilidade na cor da pele (Jablonski 2012). Esses ancestrais possuíam pele clara, macia, quase completamente coberta por pelos escuros e espessos e poucas glândulas sudoríparas. Durante a evolução, os humanos modernos adquiriram pele mais firme e resistente, com pelos mais finos, glândulas sudoríparas espalhadas por todo o corpo e pele mais escura. A redução do número de pelos e o aumento do número de glândulas sudoríparas solucionou o problema da termorregulação dos humanos no quente ambiente em que estavam, mas deixou a pele exposta aos danos solares, o que foi superado com o escurecimento da pele (Jablonski 2004, Jablonski and Chaplin 2010, Jablonski 2012). Quando os humanos se dispersaram por ambientes com menor incidência de radiação ultravioleta (UV), pele escura deixou de ser vantagem, pois essa característica dificulta a produção de vitamina D (Jablonski 2012). Assim, mutações que ampliavam as possibilidades de adaptação a esses novos ambientes foram mantidas e, em vista disso, podemos observar a enorme diversidade de cores atual.

1.1 Biologia e Genética da Pigmentação Humana

As características da pigmentação humana, tais como variações na cor dos cabelos, olhos e pele, pertencem à categoria dos mais visíveis e diferenciáveis fenótipos humanos. A pigmentação humana é primariamente explicada pela presença de melanina na epiderme, íris e cabelos e sua variação fenotípica é determinada pela quantidade, tipo e distribuição da melanina pelos melanócitos, células especializadas na síntese e distribuição deste biopolímero (Hearing 2011, Kawakami and Fisher 2011, Liu, Wen et al. 2013). Os melanócitos derivam de células precursoras (melanoblastos) durante o desenvolvimento embrionário. Os melanoblastos são originários da crista neural e, após seu desenvolvimento, migram para locais apropriados e se diferenciam em melanócitos para realizar a função de produzir padrões normais de pigmentação. Na pele humana, os melanócitos estão localizados na junção entre a derme e a epiderme num padrão característico de dispersão. Cada melanócito na camada basal da epiderme está funcionalmente conectado aos fibroblastos subjacentes na derme e aos queratinócitos sobrejacentes na epiderme (FIGURA 1). Estes três tipos de células são altamente interativos e comunicam-se entre si através de fatores de secreção e

seus receptores e através do contato célula-célula para regular processos que acarretarão no fenótipo da pele (Yamaguchi, Brenner et al. 2007, Cichorek, Wachulski et al. 2013, Duval, Cohen et al. 2014).

Não é o número, distribuição corporal ou densidade dos melanócitos em si, mas sim a regulação do processo de melanogênese que deve ser examinada para um melhor entendimento da gama dos diferentes fenótipos da pigmentação. A biogênese da melanina ocorre em organelas chamadas melanossomos, compartimentos subcelulares especializados do complexo de Golgi. Estes são transferidos dos melanócitos para os queratinócitos adjacentes na pele e nos cabelos. Enzimas melanocíticas especializadas e proteínas estruturais são transferidas e agregadas à partícula melanossomal no processo de maturação que passa por quatro estágios (Yamaguchi, Brenner et al. 2007, Sturm 2009).

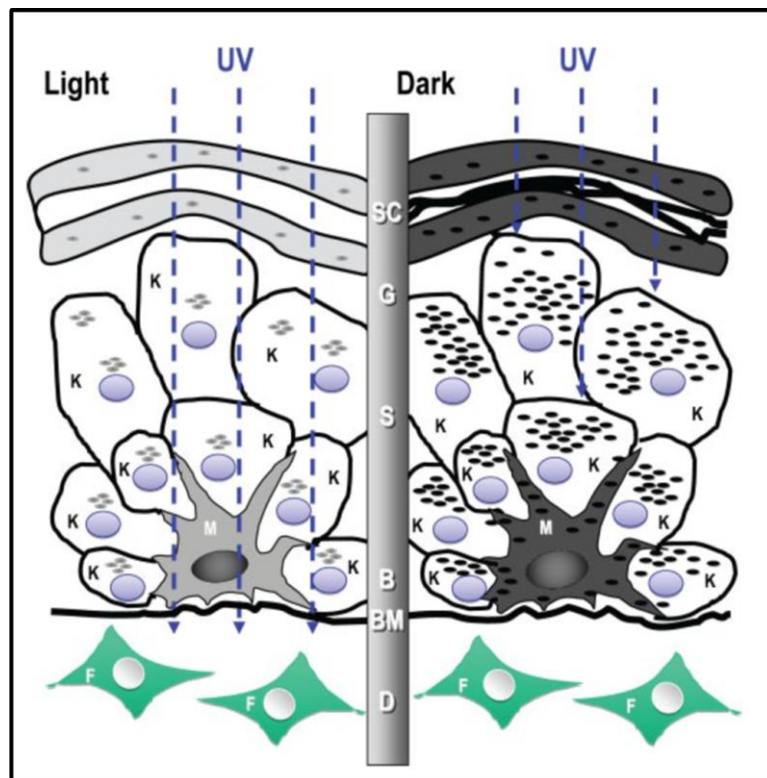


FIGURA 1. Esquema da estrutura da pele humana: tipos de pigmentação clara (*light*) e escura (*dark*). De cima para baixo: UV = radiação ultravioleta; SC = estrato córneo; G = estrato granuloso; S = estrato espinhoso; B = estrato basal; BM = membrana basal; D = derme. Tipos celulares: K = queratinócitos; M = melanócitos; F = fibroblastos; grânulos ovais = melanossomos contendo melanina.
Fonte: Yamaguchi, et al. 2007.

No primeiro estágio, os melanossomos são vacúolos vazios que brotam do retículo endoplasmático e não conseguem sintetizar melanina. Somente a partir do segundo estágio, quando proteínas estruturais e enzimáticas são transportadas para o melanossomo é que este torna-se capaz de começar a sintetizar melanina. E só

quando está totalmente maduro (quarto estágio), consistindo em uma partícula escura preenchida pela melanina que produziu durante sua maturação, é que ele é transportado para os queratinócitos da pele e folículos capilares. Nos humanos, duas formas independentes de melanossomos que diferem em tamanho, formato, estrutura e composição química já foram descritos. O eumelanossomo é uma partícula de formato oval, opaca e local da síntese da eumelanina (pigmento escuro, castanho/preto). Já a feomelanina (pigmento claro, vermelho/amarelo) deriva de etapas conjugativas envolvendo grupos sulfidríla e cisteínas que ocorrem nos feomelanossomos esféricos (Sturm 2006, Costin and Hearing 2007, Paterson, Fielder et al. 2015). A forma, o conteúdo, a transferência e o acúmulo desses melanossomos na vizinhança dos queratinócitos é que determina a cor dos cabelos e da pele, ao passo que, nos olhos, os melanossomos são retidos nos melanócitos da íris. Além disso, o grau de melanização (quantidade de eumelanina e feomelanina produzida pelos melanossomos) e distribuição dos melanossomos varia quantitativamente entre indivíduos de diversas origens biogeográficas (Sturm 2006).

A evolução dessa diversidade na pigmentação da pele em humanos foi determinada por muitos fatores. No entanto, a radiação de luz ultravioleta (UVR) tem sido o principal fator seletivo influenciando a evolução da biossíntese da melanina na pele. Atualmente, novos fatores influenciando a evolução da pigmentação da pele em humanos tem surgido, e a cultura claramente reduziu a influência da ação da seleção natural neste contexto. Comportamentos culturais, tais como a utilização de protetor solar, roupas e abrigos, se tornaram mais comuns em algumas populações, o que contribuiu para a redução da exposição de um indivíduo à UVR. A exposição da pele humana a UVR resulta numa profunda alteração do metabolismo, estrutura e função das células epidermais. Estas atividades incluem a ativação dos melanócitos, aumento na produção de melanossomos, aumento no tamanho dos complexos melanossomais incorporados pelos queratinócitos e iniciação da síntese de vitamina D (Jablonski 2004, Jablonski and Chaplin 2010, Yuen and Jablonski 2010).

Na maioria dos humanos, uma exposição casual a luz do sol leva à conversão do 7-deidrocolesterol na pele para pré-vitamina D3 pelos fótons UV e subsequente isomerização dessa última para vitamina D3 em temperatura corporal. A deficiência de vitamina D3 pode afetar a saúde humana em qualquer momento da vida e muitos de seus efeitos podem levar à um aumento da taxa de mortalidade. Conforme a quantidade de melanina na pele aumenta, aumenta o tempo necessário de exposição à luz UV para síntese de pré-vitamina D3. Peles extremamente pigmentadas tornam-se pouco adaptadas para permitir que quantidades suficientes do precursor de vitamina D3 sejam sintetizadas em ambientes de baixa intensidade de UVR. Se a

duração da exposição não for suficiente para catalisar a síntese de pré-vitamina D3, os indivíduos estarão sob um risco bem maior de deficiência de vitamina D3 e de suas manifestações (raquitismo, osteomalácia, osteoporose) (Jablonski and Chaplin 2000, Yuen and Jablonski 2010). As diversas ações biológicas da vitamina D3 se iniciam através de mudanças precisas na expressão gênica mediada pelo seu receptor (receptor de vitamina D3, VDR). O VDR é um fator de transcrição e membro da família dos receptores nucleares de hormônios esteroides que regula a expressão de mais de 200 genes que interferem na sua atividade biológica. A evolução adaptativa da pigmentação da pele permite que pessoas em diferentes latitudes produzam níveis adequados de vitamina D3 para manutenção de seu metabolismo. (Cannell, Hollis et al. 2008, Pike and Meyer 2010, Yuen and Jablonski 2010, Bikle 2014).

A genética da pigmentação humana tem sido estudada por muitos anos. A descoberta de importantes genes determinantes da pigmentação foi feita por meio da combinação de abordagens genéticas, bioquímicas e celulares, mas, indubitavelmente, foi o pronto acesso a sequência completa do genoma humano e a documentação de um vasto número de SNPs em várias populações por meio dos avanços tecnológicos dos estudos de associação em larga escala (*Genome-Wide Association Studies* - GWAS) e dos métodos estatísticos de testes de associação os maiores responsáveis pela expansão do conhecimento (Sturm 2009, Sturm and Duffy 2012, Ainger, Jagirdar et al. 2017).

Muitos genes distintos direta ou indiretamente relacionados à pigmentação já foram identificados. Vários deles afetam processos críticos ao desenvolvimento dos melanoblastos (células precursoras dos melanócitos), outros regulam a diferenciação e sobrevivência dos melanócitos, alguns estão envolvidos na biogênese e funcionamento dos melanossomos e outros regulam processos distintos que influenciam a pigmentação. Por exemplo, genes que codificam enzimas críticas incluem o gene da tirosinase (*TYR*), *TYRP1* e *DCT*, cujas mutações afetam drasticamente a quantidade e qualidade da melanina sintetizada. Genes de proteínas estruturais críticas incluem *Pmel17* e *MART1*, ambas necessárias para a maturação estrutural do melanossomo. Existem ainda, genes que codificam proteínas envolvidas no processo de sinalização (*MC1R*) e proteínas envolvidas na distribuição e transporte de proteínas/íons aos melanossomos (*SLC45A2* e *SLC24A5*) (Yamaguchi, Brenner et al. 2007).

1.1.1 Impacto funcional de mutações em genes de pigmentação

Polimorfismos localizados em regiões exônicas, que causam mudanças de aminoácido alterando a proteína, podem afetar o fenótipo caso a proteína mutada seja truncada ou perca completamente a função. Além disso, para gerar um RNAm correto e maduro, os éxons devem ser identificados e unidos precisa e eficientemente. Mutações sem sentido, de sentido trocado ou até mesmo silenciosas podem inativar genes pela indução da maquinaria de *splicing* a excluir o éxon mutante formando um RNAm aberrante que é instável ou codifica uma isoforma proteica defeituosa ou deletéria (Cartegni, Chew et al. 2002). Por não causarem uma mudança no aminoácido e, conseqüentemente, não afetarem o produto proteico, SNPs localizados em regiões intrônicas sofrem uma menor pressão seletiva e, assim, podemos observar um maior número de variações nessas regiões. Porém, é importante ressaltar que polimorfismos presentes em íntrons também podem desempenhar funções de regulação da expressão gênica, por afetarem eventos de *splicing* e/ou a estabilidade do RNAm (Sun and Maquat 2000, Comeron 2004, Chamary, Parmley et al. 2006, Wang, Feng et al. 2007). Sabe-se hoje que as regiões não traduzidas do RNAm têm um papel indispensável na regulação da expressão gênica. Sua interação com proteínas de regulação e outros RNAs (por exemplo, microRNAs) podem influenciar a estabilidade, transporte, processamento, eficiência da tradução e a degradação do RNAm (Matoulkova, Michalova et al. 2012).

Estima-se que cerca de 60% dos genes codificadores em humanos sejam diretamente regulados por microRNAs (Friedman, Farh et al. 2009, Catalanotto, Cogoni et al. 2016). MicroRNAs (miRNAs) são pequenas moléculas de RNA de fita simples com 18-22 nucleotídeos de comprimento. A função dos miRNAs é inibir a síntese proteica parcial ou completamente através da sua interação com sequências de bases na 3'UTR (do inglês, 3' *untranslated region*) do RNAm regulando a expressão gênica (Kamaraj, Gopalakrishnan et al. 2014, Gulyaeva and Kushlinskiy 2016). SNPs na 3'UTR podem criar novos sítios de ligação para miRNAs ou excluir um sítio de ligação que existia naquela região, resultando em uma variedade de possíveis interações entre a sequência alvo e os miRNAs.

A transcrição de genes eucarióticos é precedida por diversos eventos como remodelamento da cromatina, modificações de histona, ligação de fatores de transcrição aos promotores e recrutamento da RNA polimerase junto ao promotor (Smale and Kadonaga 2003). Dessa forma, polimorfismos na região promotora podem afetar o processo de transcrição pela RNA polimerase ao alterar a afinidade entre as proteínas interagindo nesse processo ou recrutando novos fatores de transcrição a se

ligarem nessas regiões (Albert and Kruglyak 2015). Além disso, variantes em locais de ligação de fatores de transcrição se correlacionam ainda com modificações de histonas, níveis diferenciais de RNAm e metilação do DNA (Banovich, Lan et al. 2014).

Mutações nas diferentes regiões em genes de pigmentação podem levar aos diversos fenótipos normais observados em humanos, mas também podem ocasionar doenças, como o albinismo. Todas as formas de albinismo resultam de disfunções de enzimas e proteínas melanogênicas, o que leva a uma pigmentação da pele, cabelos e olhos comprometida. Até agora, cinco tipos de albinismo foram definidos mapeando-se cinco diferentes *loci* relacionados à pigmentação. Mutações nos genes *TYR* (albinismo oculocutâneo tipo 1 - OCA1) e *TYRP1* (OCA3) podem conduzir a baixos níveis de produção de tirosinase e à degradação proteossomal da mesma. Mutações nos genes *P* (OCA2) e *SLC45A2* (OCA4) podem influenciar o transporte e distribuição de tirosinase funcional para os melanossomos. Por fim, mutações no gene *GPR143* (*G protein-coupled receptor 143*), anteriormente conhecido como *OA1* (albinismo ocular tipo 1), podem comprometer a biogênese do melanossomo e a pigmentação através de um mecanismo ainda desconhecido e, dessa forma, perturbar a produção de melanina (Yamaguchi, Brenner et al. 2007, Yamaguchi and Hearing 2014).

Assim sendo, pode-se observar o quão complexa é a pigmentação. Inúmeros receptores, ligantes e outros fatores atuam de forma complexa na regulação da pigmentação, constituindo vias que devem funcionar perfeitamente para não ocorrerem danos no processo. Deste modo, existem inúmeras possibilidades de mutações que podem não necessariamente causar danos, mas sim a variação normal nas cores da pele, olhos e cabelos.

1.2 *SLC45A2*

O gene *SLC45A2* (*solute carrier family 45, member 2*) codifica a proteína MATP (*membrane associated transport protein*) que está envolvida na síntese de melanina através do transporte de prótons via membrana melanossomal e, conseqüentemente, controle do pH e da homeostase iônica dentro do melanossomo. Adicionalmente, atua também no transporte intracelular da tirosinase por meio de um mecanismo ainda não identificado (Tully 2007, Dooley, Schwarz et al. 2013, Kondo, Namiki et al. 2015). O gene *SLC45A2* está localizado na região cromossômica 5p13.2 e tem 7 éxons abrangendo pouco mais de 40 Kb (Graf, Hodgson et al. 2005).

A proteína MATP se distribui ao longo da bicamada lipídica da membrana melanossomal, assim como a proteína P associada ao OCA2 (Costin, Valencia et al. 2003). Existe a hipótese de que a proteína P seja uma transportadora de ânions na

membrana melanossomal. Se a MATP está similarmente localizada na membrana e transporta prótons para controle do pH melanossomal, então a ação dessas duas proteínas pode resultar numa acidificação normal do melanossomo (Newton, Cohen-Barak et al. 2001).

Como mencionado, a proteína MATP também está relacionada ao transporte de tirosinase (TYR) através de um mecanismo ainda desconhecido (Kondo, Namiki et al. 2015). A enzima TYR é responsável pela conversão da tirosina em L-DOPA no primeiro passo da via bioquímica da produção da melanina. O pH melanossomal é um fator crucial na síntese da melanina em dois aspectos: (a) nos melanossomos em estágio inicial de maturação, o pH ácido contribui para estabilizar a L-DOPA prevenindo a auto-oxidação e (b) durante a maturação dos melanossomos, o aumento gradual do pH otimiza a atividade da tirosinase para a produção de melanina (Bin, Bhin et al. 2015). Assim, o pH intracelular é essencial na regulação da função da tirosinase, não somente porque o pH intramelanossomal afeta drasticamente as funções catalíticas, mas também porque um correto gradiente de pH é fundamental na via responsável pela distribuição das proteínas melanossomais (Yamaguchi, Brenner et al. 2007).

Sob condições normais, a MATP aumenta o pH melanossomal funcionando como um transportador de prótons. Deste modo, o cobre pode se ligar a tirosinase e torná-la ativa. Quando mutada, MATP é responsável pelo albinismo oculocutâneo tipo 4 (OCA4) (Newton, Cohen-Barak et al. 2001, Kondo, Namiki et al. 2015). Nos melanossomos dos indivíduos afetados por OCA4, a MATP não funciona normalmente e o lúmen melanossomal se torna ácido. Assim, o cobre não consegue ser incorporado pela TYR, reduzindo a atividade da mesma (FIGURA 2) (Bin, Bhin et al. 2015).

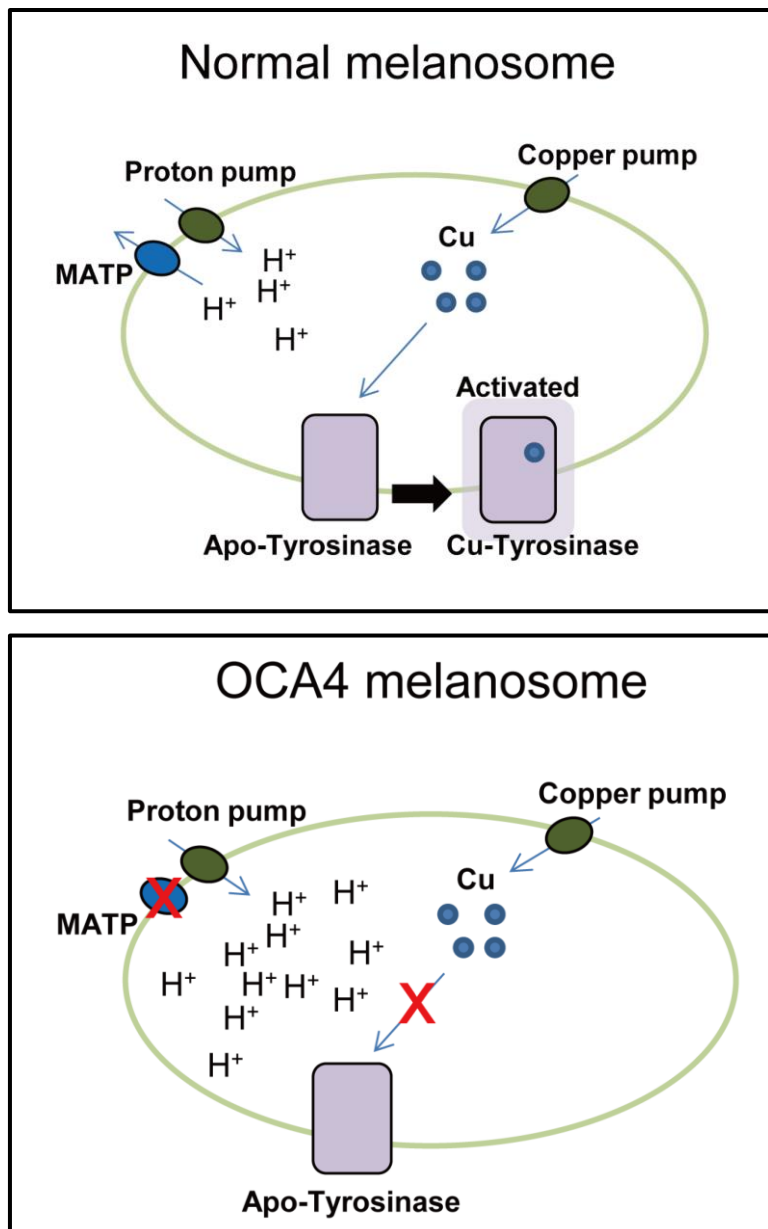


FIGURA 2. Esquema da função transportadora da proteína MATP na regulação do pH melanosomal em situação normal e na ocorrência de OCA4. *Proton pump* = bomba de prótons; *Copper pump* = bomba de cobre; Cu = cobre; *Apo-Tyrosinase* = tirosinase sem cobre (atividade reduzida); *Cu-Tyrosinase* = tirosinase com cobre (ativada).

Fonte: Bin, et al. 2015.

Muitas variantes encontradas nesse gene podem causar sérios problemas pigmentares (incluindo o OCA4 e melanomas) (Fernandez, Milne et al. 2008, Ibarrola-Villava, Fernandez et al. 2011, Kamaraj and Purohit 2014) mas, ao mesmo tempo, já foi descrito o envolvimento de seus polimorfismos na variação normal da pigmentação (Kayser 2015, Siewierska-Gorska, Sitek et al. 2017). Desde sua descoberta em 2001 (Newton, Cohen-Barak et al. 2001), muitos trabalhos associaram suas variantes com as cores de olhos, cabelos e pele em humanos. Em nível populacional, foi relatado que o alelo Leu374 (rs16891982*C) do SNP rs16891982 está presente com uma

frequência significativamente maior em asiáticos (0,887), afro-americanos (0,586) e aborígenes australianos (0,725) do que em caucasianos (0,066). Caucasianos com esse alelo tendem a ter pigmentação mais escura, estando as maiores frequências em indivíduos com cabelos pretos [*Odds Ratio* (OR) = 25,6], pele escura (OR = 28,6) e olhos escuros (OR = 3,5) (Graf, Hodgson et al. 2005). Outro trabalho (Yuasa, Umetsu et al. 2004) corroborou tal observação ao revelar que o alelo alternativo Phe374 estava presente em alta frequência (0,96) em alemães, enquanto completamente ausente em japoneses. O Projeto 1000 Genomas também mostra uma frequência do alelo Phe374 muito maior em europeus (95,9% em finlandeses e 97,8% em britânicos) do que em outras populações estudadas (21,3% em afro-americanos, 8,7% em indianos, 1,5% em chineses e completa ausência em nigerianos) (<http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/> 2018).

Em trabalho realizado mais recentemente (Beleza, Santos et al. 2013), foi observado que a seleção favorecendo o alelo rs16891982*G (374Phe) teve início estimado entre 11 mil e 19 mil anos atrás, bem depois da expansão dos humanos modernos para fora da África, o que sugere que esse evento de seleção pode ter sido influenciado pelo aumento das populações europeias, o que teria diminuído a intensidade da deriva e favorecido o acúmulo de mutações vantajosas em um relativamente curto período de tempo. O período que compreendeu a seleção desse alelo coincide com o último máximo glacial (Clark, Dyke et al. 2009), o que torna possível que as condições climáticas da época tenham favorecido mudanças culturais que culminassem na redução de exposição ao sol e na procura por abrigo e uso de roupas para proteção, aumentando o risco de insuficiência de vitamina D (Jablonski and Chaplin 2000) e, dessa forma, selecionando de forma positiva a pigmentação mais clara.

Tanto esse SNP quanto outro intrônico do gene *SLC45A2* (rs28777) fazem parte do sistema HRISplex. Este sistema, composto por 24 variantes presentes em genes de pigmentação, foi desenvolvido para a predição das cores dos olhos e cabelos para aplicações forenses e antropológicas, e possibilita a predição das cores de olhos e cabelos em humanos com uma acurácia maior do que 74% (Walsh, Liu et al. 2013). Em uma pesquisa sobre predição de cores de cabelos, o alelo C do SNP rs28777 se mostrou associado a cabelos pretos [OR = 7,05 (2,2 – 22,3)] em poloneses (Branicki, Liu et al. 2011), enquanto o alelo A é encontrado em menor frequência (17,68%) em nigerianos e maior frequência (98,48%) em euro-americanos, caracteristicamente conhecidos por terem fenótipos pigmentares mais claros (<http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/> 2018).

Quando observamos que mais de 4300 variantes já foram identificadas, podemos considerar que o gene *SLC45A2* é bem diverso, (<https://www.ncbi.nlm.nih.gov/gene/51151>, acesso em 18/05/2018). Dentre elas, a grande maioria é composta por SNPs (3794), mas destes apenas 34 foram relacionados em publicações. Trabalhos recentes continuam a comprovar a importância do gene *SLC45A2* na pigmentação humana, tanto em termos de descoberta de novas mutações associadas ao OCA4 (Simeonov, Wang et al. 2013, Wei, Zang et al. 2015, Toth, Fabos et al. 2017) ou predisposição ao melanoma (Fesenko, Chudinov et al. 2016, Park, Talukder et al. 2017), sendo útil até mesmo no diagnóstico pré-natal, quanto nas descobertas de mutações relacionadas à variação normal e predição da pigmentação humana (Beleza, Johnson et al. 2013, Liu, Visser et al. 2015, Fracasso, de Andrade et al. 2017, Siewierska-Gorska, Sitek et al. 2017).

1.3 TYR

O gene *TYR* (*tyrosinase*) está localizado na região cromossômica 11q14.3. Este gene possui 117.888pb, 5 éxons e codifica a enzima tirosinase composta por 529 aminoácidos. Tal proteína catalisa os dois primeiros passos da conversão de tirosina em melanina, além de atuar em uma etapa final da biossíntese de eumelanina. Quando mutada é responsável pelo albinismo oculocutâneo tipo 1 (OCA1) (Bae-Harboe and Park 2012).

A observação de que a tirosinase se localiza em vesículas do complexo de Golgi e na membrana dos melanossomos suportava a hipótese de que moléculas de tirosinase se agregam ou combinam com outros componentes como proteínas melanossomais para formar estruturas granulosas saindo do complexo de Golgi para os melanossomos (Yamamoto and Takeuchi 1981). Trabalhos posteriores confirmaram essa hipótese, ampliando o conhecimento não somente sobre sua localização mas também sobre mecanismos envolvidos na melanogênese (Orlow 1995, Calvo, Frank et al. 1999, Theos, Tenza et al. 2005). A tirosinase, assim como a maioria das proteínas constituintes dos melanossomos nos estágios II-IV são glicoproteínas integrais de membrana, sintetizadas pelos ribossomos associados ao retículo endoplasmático. De modo geral, se tornam integradas à membrana do retículo endoplasmático, são glicosiladas por um complexo proteico e são posteriormente dobradas e montadas por proteínas e enzimas do próprio retículo. O fracasso no dobramento correto dessas enzimas, como ocorre em várias formas de albinismo, resulta no extravio e na liberação anormal dessas proteínas no meio, e em posterior degradação. Algumas evidências sugerem que os genes *TYRP1*, *OCA2* e *SLC45A2*

regulam o dobramento e a saída da TYR do retículo endoplasmático por mecanismos ainda pouco conhecidos (Costin, Valencia et al. 2003, Sitaram and Marks 2012, Dooley, Schwarz et al. 2013).

Como mencionado previamente, a maturação dos melanossomos do estágio II para os estágios III e IV requer a presença de enzimas melanossomais e transportadoras, como a TYR. A tirosinase possui a maior parte de seu domínio no lúmen, um domínio transmembrana com 26 aminoácidos e um domínio citoplasmático com 30 aminoácidos. A conversão da tirosina em melanina ocorre no lúmen. A descoberta de que a tirosinase regula três passos importantes na melanogênese, com a habilidade de catalisar essas três reações separadamente em uma única via, ocorreu no início dos anos 80 (Korner and Pawelek 1982). Entretanto, o preciso mecanismo catalítico dessa conversão, bem como a interação de substratos com os dois sítios de ativação de íons cobre, ainda é um importante tópico de investigação (Lai, Wichers et al. 2018). O domínio citoplasmático da TYR atua no reconhecimento de outras proteínas de membrana e/ou como sinais para endocitose. Tal domínio interage com sítios conservados em complexos adaptadores proteicos (AP). Os APs pertencem a uma família de cinco complexos relacionados que funcionam como adaptadores de reconhecimento. Eles são recrutados para a membrana e se ligam às proteínas carregadoras integrais de membrana que contêm sítios de ligação de AP, agrupando-se em domínios de revestimento de membrana. A membrana se dobra formando vesículas que são direcionadas para outras membranas alvo, por exemplo, para as membranas melanossomais. Interações no domínio citoplasmático da TYR com APs podem ajudar no seu direcionamento através dessas vesículas e de túbulos para as membranas melanossomais (FIGURA 3) (Theos, Tenza et al. 2005, Sitaram and Marks 2012, Jani, Purushothaman et al. 2015).

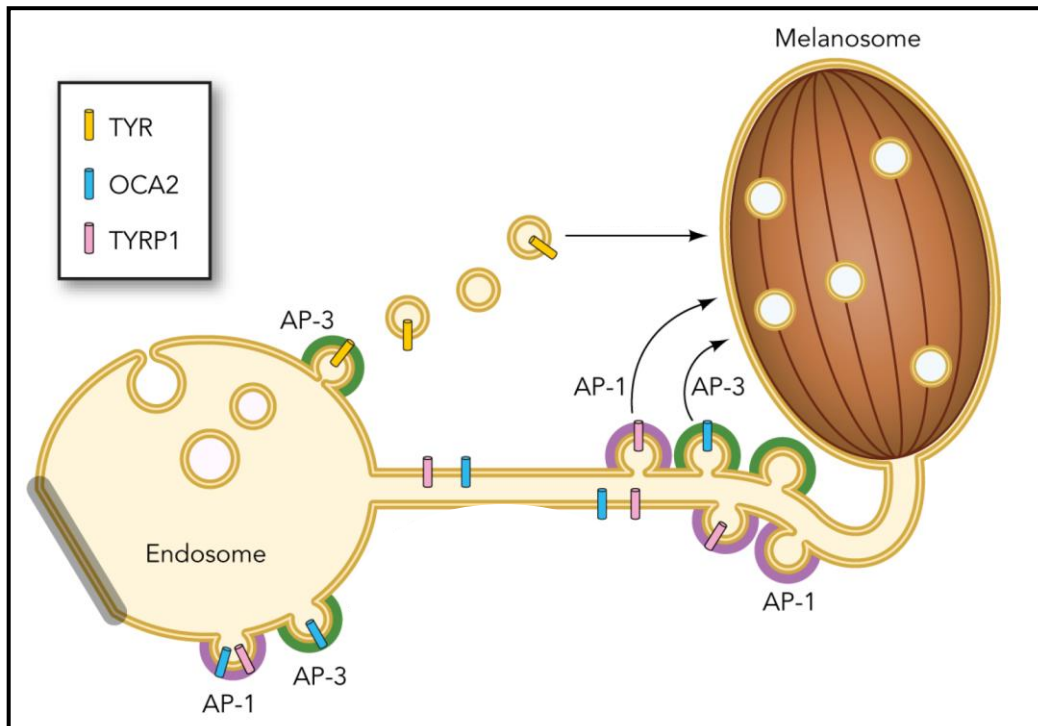


FIGURA 3. Esquema da interação entre as proteínas de membrana (TYR, TYRP1 e OCA2) com adaptadores proteicos (AP) e posterior transporte para as membranas melanossomais através de vesículas isoladas (no caso da TYR) ou de túbulos que se originam nos endossomos e se fundem com a membrana de melanossomos próximos.

Fonte: Adaptada de Sitaram, et al. 2012.

Polimorfismos neste e nos outros domínios da TYR já foram associados com doenças pigmentares como OCA1 e queratose actínica, assim como com fenótipos normais de pigmentação (Cerqueira, Hunemeier et al. 2014, Jacobs, Liu et al. 2015, Kayser 2015, Mondal, Sengupta et al. 2016). Dois de seus SNPs, rs1393350 e rs1042602, fazem parte do sistema HlrisPlex (Walsh, Liu et al. 2013). O alelo A do SNP rs1393350 foi associado a cabelos loiros (OR = 1,36 [1,12-1,66]; $p = 0,00011$), olhos azuis (OR = 1,43 [1,21-1,71]; $p = 3,3 \times 10^{-12}$) e sardas (OR = 1,13 [1,00-1,28]; $p = 0,0029$) em estudo GWAS (Sulem, Gudbjartsson et al. 2007). Porém, em um trabalho posterior (Branicki, Liu et al. 2011), o mesmo alelo se mostrou associado com cabelos castanhos (OR = 1,70 [1,02-2,82]; $p = 0,041$). Com relação ao SNP rs1042602 (Stokowski, Pant et al. 2007), o alelo C se mostrou associado a pele escura (OR = 4,36 [2,64-7,2]; $p = 4,48 \times 10^{-10}$), o que concordou com dados mais recentes de que o alelo alternativo (rs1042602*A) deste mesmo SNP estava associado com fenótipos de pigmentação claros em australianos (Jagirdar, Smit et al. 2014). Em uma pesquisa mais atual desenvolvida na Índia, o alelo A desse SNP (rs1042602) estava presente em baixa frequência (3,4% a 11,7%) entre as seis populações estudadas naquela região, enquanto que o alelo C, associado com peles mais escuras (característico de populações da Índia) foi encontrado com uma frequência maior (Jonnalagadda, Norton

et al. 2016). Os dados destes trabalhos estão de acordo com as frequências encontradas pelo Projeto 1000 Genomas em que o alelo A pode ser observado com uma frequência baixa variando de 0% a 11,2% nas populações sul Asiáticas e bem mais frequente em europeus (ingleses 35,7% e espanhóis 39,2%) e em euro-americanos (população CEU, residentes de Utah com ancestralidade europeia, - 40%) (<http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/> 2018). O surgimento do alelo rs1042602*A foi datado de 6 a 15 mil anos atrás. Embora o aparecimento de pressões seletivas que possam ter afetado sua frequência não tenham sido detectadas no estudo em questão, é possível que vários mecanismos evolutivos tenham afetado a diversidade genética dentro dos diferentes *loci* de pigmentação naquele período: uma fraca seleção balanceadora pode ter sido responsável pelo excesso de alelos não-sinônimos específicos em europeus no gene *TYR*, enquanto que uma forte seleção positiva levou à quase completa fixação de substituições com troca de aminoácidos no gene *SLC45A2* (Hudjashov, Villems et al. 2013). O gene *TYR* possui enorme variabilidade (<https://www.ncbi.nlm.nih.gov/gene/7299>, acesso em 18/05/2018), porém muitos de seus polimorfismos já caracterizados falharam em mostrar associações significativas com alguma pigmentação de pele específica (Dooley, Schwarz et al. 2013). Dentre as 13.290 variantes já relacionadas no banco de dados do NCBI (Geer, Marchler-Bauer et al. 2010), mais de 11.900 são SNPs. Dentre o número reduzido de publicações com alguns desses SNPs, podemos citar achados como o SNP rs1126809*G associado com a cor da pele escura na população brasileira (Cerqueira, Hunemeier et al. 2014), o rs1393350 associado com a cor dos cabelos em euro-americanos e euro-australianos (Han, Kraft et al. 2008) e o rs12421746*T associado com a cor de cabelos loiros em dinamarqueses (Mengel-From, Wong et al. 2009).

Apesar da tirosinase ser uma das oxidases multicobre mais bem estudadas, de sua atividade de conversão de tirosina em melanina ter sido descoberta a mais de um século e de mais de 200 mutações terem sido associadas com albinismo (Solomon, Sundaram et al. 1996, Oetting 2000, Mondal, Sengupta et al. 2016), muitas informações sobre sua estrutura e função, os processos bioquímicos que levam ao transporte adequado dessa enzima para os melanossomos e como estes fatores e a diversidade de polimorfismos em seu gene influenciam a associação de alelos com a pigmentação normal ainda precisam ser elucidados (Ramsden and Riley 2014, Kondo, Namiki et al. 2015, Lai, Wichers et al. 2018).

A despeito do conhecimento de que a proteína codificada pelo gene *SLC45A2* está envolvida no transporte e distribuição de tirosinase funcional para os melanossomos, o mecanismo pelo qual ambas proteínas interagem e como isso pode

influenciar a determinação das cores de cabelos, olhos e pele em humanos ainda não foi descoberto (Kondo, Namiki et al. 2015), todavia, avaliar a diversidade genética desses dois genes (*SLC45A2* e *TYR*) em uma população miscigenada como a brasileira pode expandir nosso conhecimento sobre a pigmentação humana.

1.4 População Brasileira

A população brasileira é uma das mais heterogêneas em todo o mundo, resultado de cinco séculos de mistura inter-étnica envolvendo principalmente Nativos Americanos, Europeus e Africanos (Salzano and Sans 2014). Estima-se que antes da chegada dos colonizadores, por volta de 2,5 milhões de indígenas já habitavam o território brasileiro (Salzano and Bortolini 2002). Os nativos sofreram então uma drástica redução populacional tanto devido aos conflitos com os colonizadores quanto às doenças trazidas por eles, tais como varíola e sarampo (IBGE 2007). Nos três primeiros séculos de colonização, aproximadamente 500.000 indivíduos vieram de Portugal (Santos, Ribeiro-Rodrigues et al. 2010). Entre 1820 e 1975, a grande maioria dos 6 milhões de imigrantes europeus que chegaram ao Brasil era de origem predominantemente portuguesa e italiana (70% do total), tendo ocorrido também a entrada de espanhóis, alemães, sírios e japoneses. Entre os séculos XVI e XIX, aproximadamente 3,5 milhões de africanos foram trazidos como escravos para o Brasil, vindos principalmente das regiões Oeste, Centro-Oeste e Sudeste da África (Santos and Coimbra Jr. 1994).

Isso pode ser verificado nos resultados do estudo realizado por Pena et al. (2011) em que a maior contribuição relativa de ancestralidade observada nos brasileiros é predominantemente europeia (com as proporções variando de 60,6% no Nordeste até 77,7% no Sul do Brasil), seguida de africana (30,3% no Nordeste a 10,9% no Norte) e ameríndia (19,4% no Norte a aproximadamente 9% nas outras três regiões) (Pena, Di Pietro et al. 2011). Em um trabalho mais recente, as proporções de ancestralidade europeia, africana e ameríndia para as regiões Sudeste (78.5% europeia, 14.7% africana e 6.7% ameríndia) e Sul (76.1% europeia, 15.9% africana e 8% ameríndia) foram muito parecidas com o trabalho de Pena et al., mas um pouco diferentes no Nordeste (42.9% europeia, 50.8% africana e 6.4% ameríndia) (Kehdy, Gouveia et al. 2015), não obstante indicando ampla contribuição europeia nas regiões analisadas.

De acordo com uma estimativa realizada em 2017 baseada no censo de 2010 (IBGE 2010), o Brasil possui mais de 200 milhões de habitantes, dos quais 47,7% se autodeclararam "brancos", 43,1% "pardos", 7,6% "negros" e 1,5% disseram pertencer a outras categorias. Estas proporções variam consideravelmente nas diferentes

regiões geográficas brasileiras devido aos aspectos particulares de suas colonizações: houve uma contribuição mais pronunciada dos nativos americanos no Norte, uma alta contribuição africana no Nordeste e uma contribuição relativamente baixa de ambas no Sul (Salzano and Sans 2014). No estado de São Paulo (com aproximadamente 45 milhões de habitantes), as proporções foram de 63,9% autodeclarados “brancos”, 29,1% “pardos”, 5,5% “negros” e 1,5% de outras categorias. Na cidade de Ribeirão Preto, que conta com 682.302 habitantes, as frequências de “brancos”, “pardos” e “negros” relatadas foram de 69,8%, 22,8% e 6,4% respectivamente (IBGE 2010). Estudos baseados em marcadores genéticos mostraram que a população urbana de Ribeirão Preto é altamente miscigenada, sendo os brancos caracterizados por 79% de contribuição europeia, 14% de africana e 7% ameríndia (Ferreira, Mendes-Junior et al. 2006); os mulatos, por 62% de contribuição europeia, 26% de africana e 12% de ameríndia, e os negros, basicamente por contribuições europeia (37%) e africana (63%) (Muniz, Ferreira et al. 2008).

Todos estes dados demonstram que a população brasileira possui uma elevada taxa de miscigenação, representada pelas contribuições de indivíduos de origens étnicas distintas. Tal heterogeneidade étnica se reflete na alta diversidade fenotípica vista nos brasileiros e dessa forma, podemos considerar que a população brasileira apresenta condições propícias para a identificação de novos alelos e haplótipos em genes de pigmentação e sua possível relação com fenótipos morfológicos (Marano 2011).

2. HIPÓTESE

Existe uma associação entre a intensidade de pigmentação dos olhos, pele e cabelos em populações urbanas brasileiras e os alelos e haplótipos compostos por polimorfismos localizados nas regiões codificadoras e regulatórias (promotora e 3'UTR) dos genes *SLC45A2* e *TYR* devido à uma possível alteração na conformação proteica e afinidade entre os diferentes alelos e haplótipos das regiões promotora e 3'UTR com fatores de transcrição e microRNAs, respectivamente.

3. JUSTIFICATIVA

Somente recentemente, mais informações sobre a função do gene *SLC45A2* nos melanócitos humanos foram descobertas e ainda pouco se sabe sobre sua interação com outros genes, embora mutações nesse gene tenham sido descobertas há mais de uma década (Cullinane, Vilboux et al. 2011, Bin, Bhin et al. 2015, Kondo, Namiki et al. 2015). Da mesma forma, muitas mutações no gene *TYR* já foram associadas com

albinismo (Mondal, Sengupta et al. 2016), porém diversas questões sobre a função, estrutura e transporte da tirosinase envolvidos com fenótipos normais de pigmentação ainda permanecem sem total compreensão (Kondo, Namiki et al. 2015, Lai, Wichers et al. 2018). Em vista disso, estudos aprofundados são necessários para o melhor entendimento das interações entre variantes alélicas em genes de pigmentação e moléculas regulatórias. E assim, o conhecimento sobre a diversidade genética desses dois genes (*SLC45A2* e *TYR*) em uma população de natureza tri-híbrida como a brasileira seria de grande valia no entendimento da genética por trás da pigmentação humana.

A predição de fenótipos complexos humanos a partir de genótipos é uma importante questão, tanto na genética médica quanto na genética forense. No caso da genética forense, a predição de características fenotípicas como a pigmentação humana pode constituir um importante fator nos casos em que a polícia não tenha pistas sobre quem teria deixado a amostra na cena do crime e o perfil de DNA convencional não tenha fornecido informações úteis para encontrar o criminoso (Liu, Wen et al. 2013).

Sendo assim, informações advindas das análises dos polimorfismos de genes como o *SLC45A2* e *TYR*, seriam de grande importância na medicina personalizada (distúrbios relacionados a hiper e hipopigmentação), genética forense e em estudos sobre adaptação, seleção e evolução humana.

4. OBJETIVO

O presente trabalho tem como objetivo principal analisar, através do sequenciamento de nova geração, a diversidade genética das regiões codificantes, promotora e 3'UTR de dois dos principais genes envolvidos na biossíntese de melanina (*SLC45A2* e *TYR*) em amostras de 340 indivíduos da população do Estado de São Paulo. Dessa forma, será possível:

- Identificar novos polimorfismos em cada um destes genes;
- Determinar as frequências de alelos, genótipos e haplótipos compostos pelos SNPs de cada região gênica analisada;
- Identificar marcadores genéticos associados à pigmentação da pele, olhos e cabelos e à presença de sardas;
- Avaliar a intensidade do desequilíbrio de ligação (LD) entre as regiões analisadas em cada gene;

- Observar se existe uma correlação entre haplótipos destes dois genes e a intensidade da pigmentação;
- Avaliar computacionalmente a interação de TFs e microRNAs com respectivos polimorfismos presentes nas regiões promotora e 3'UTR dos dois genes.
 - Identificar um conjunto de polimorfismos funcionais de cada gene que contribuam para a determinação da pigmentação da pele, cabelos e olhos, e presença de sardas, em indivíduos da população brasileira.

5. MATERIAL E MÉTODOS

5.1 Delineamento Experimental

Para elaboração deste trabalho, foi extraído o DNA de amostras de sangue de 340 indivíduos e posteriormente foi feita preparação das bibliotecas para o sequenciamento. Um ensaio customizado foi realizado com o kit de enriquecimento de alvo *HaloPlex* (*Agilent Technologies*) utilizando-se a ferramenta *SureDesign* (*Agilent Technologies* - disponível em <https://earray.chem.agilent.com/suredesign/>). Sondas foram geradas de modo a cobrir as regiões promotora, 5'UTR, CDS (*Coding DNA Sequence*), regiões intrônicas flanqueadoras de cada um dos éxons e a região 3'UTR dos genes *SLC45A2* e *TYR* (FIGURAS 4 e 5). A ferramenta utilizou o release hg19/GRCh37 do genoma humano como referência para a construção das sondas que abrangeram cada um dos genes.

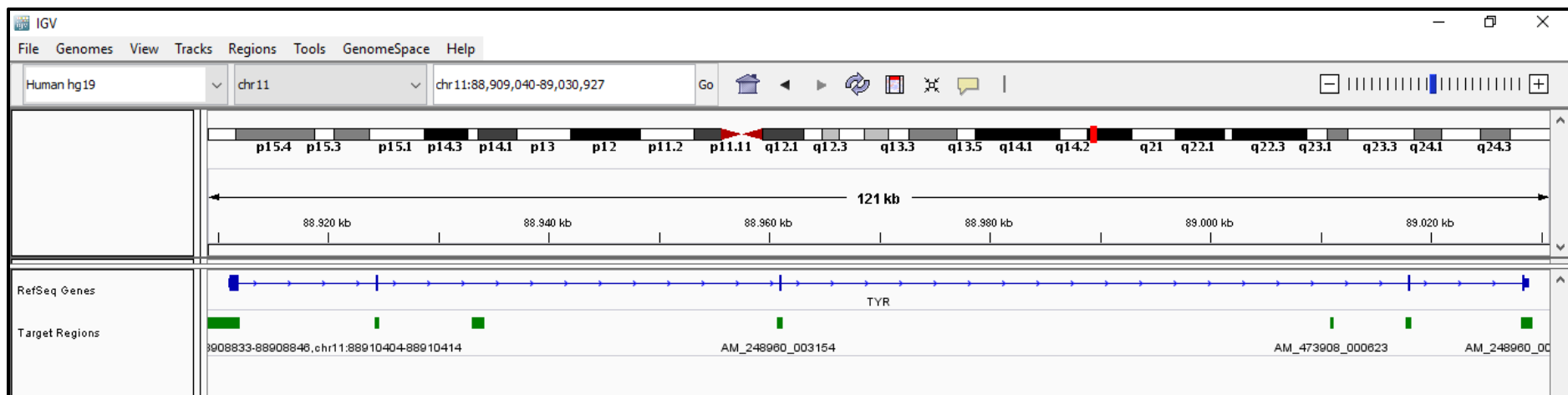


FIGURA 4. Representação do gene *TYR* no programa IGV_2.3.66. A parte principal da tela de visualização é dividida em 2 partes. No topo pode ser vista a representação esquemática do cromossomo 11 e a região deste cromossomo que está sendo visualizada (indicada por uma barra vermelha). Logo abaixo, é apresentada uma escala contendo as coordenadas da região ampliada. Na parte inferior da tela temos a anotação do genoma para a região visualizada (*RefSeq Genes*) em azul e em verde, a anotação das regiões cobertas pelo nosso ensaio.

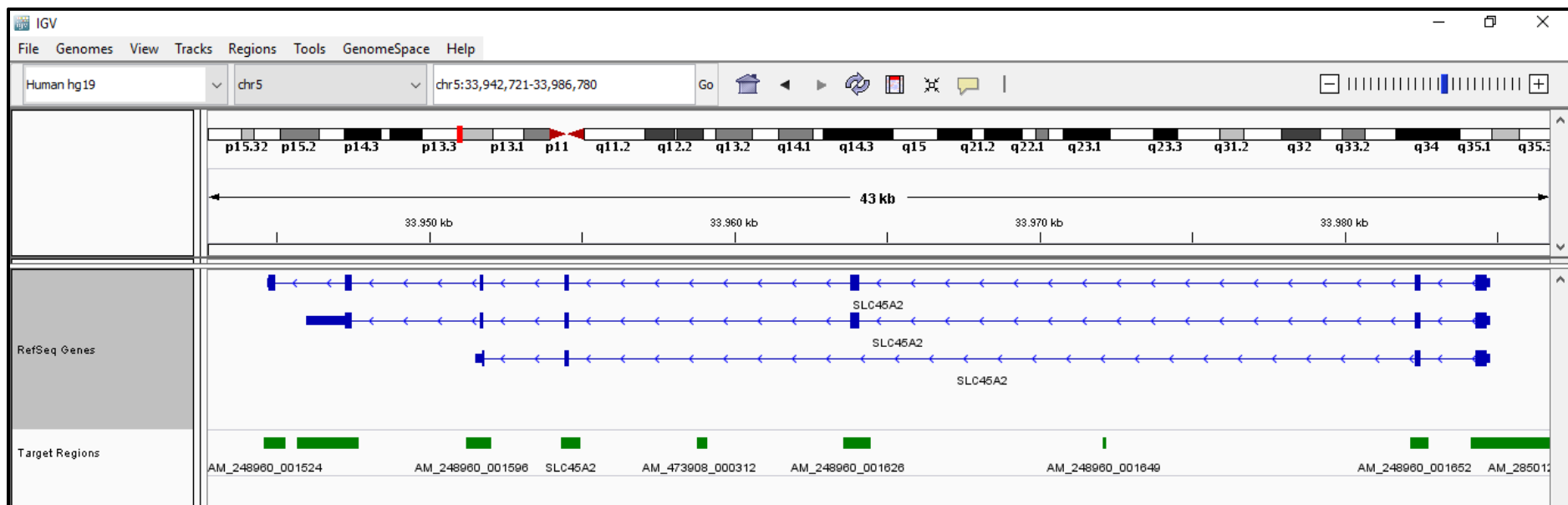


FIGURA 5. Representação do gene *SLC45A2* no programa IGV 2.3.66. A parte principal da tela de visualização é dividida em 2 partes. No topo pode ser vista a representação esquemática do cromossomo 5 e a região deste cromossomo que está sendo visualizada (indicada por uma barra vermelha). Logo abaixo, é apresentada uma escala contendo as coordenadas da região ampliada. Na parte inferior da tela temos a anotação do genoma para a região visualizada (*RefSeq Genes*) em azul com as três isoformas do gene analisadas neste trabalho (acima a isoforma com sete éxons, no meio a isoforma com seis éxons e abaixo a isoforma com quatro éxons). Em verde, a anotação das regiões cobertas pelo nosso ensaio.

Após o sequenciamento, as etapas seguintes de análises podem ser sumarizadas em duas fases principais: (1) a fase de pré-processamento e chamada de variantes, que consistiu na geração dos dados no formato FastQ, aplicação de filtros de qualidade, alinhamento a um genoma de referência e determinação das variantes através de comparação das sequências alvo a um genoma de referência e (2) a fase final de análises estatísticas, que incluiu o faseamento computacional dos haplótipos encontrados, avaliação do desequilíbrio de ligação entre os *loci* estudados, verificação das frequências alélicas, aderência ao equilíbrio de Hardy-Weinberg e associação com as características fenotípicas.

5.2 Amostra Populacional

O projeto de pesquisa do presente trabalho, juntamente com o questionário sobre características étnicas e morfológicas foi aprovado pela Comissão de Ética em Pesquisa da Instituição (FFCLRP-USP) de acordo com o processo CAAE nº 25696413.7.0000.5407 (ANEXO C).

Foram coletadas amostras de sangue (10ml) de um total de 340 voluntários não aparentados dos sexos feminino e masculino, com idade variando de 18 a 70 anos, recrutados, em sua grande maioria, no Hemocentro de Ribeirão Preto (HCFMRP-USP e Posto de Coleta - Centro). Os indivíduos que não atendiam os critérios de doação de sangue para o hemocentro mas que possuíam fenótipos raros de interesse, foram convidados a doar sangue apenas para a pesquisa nas dependências do Laboratório de Genética Bioquímica (Bloco B - Genética). Por meio de questionário, os indivíduos foram classificados em grupos de acordo com a cor dos olhos, cabelos e pele, bem como quanto à presença ou ausência de eféides (sardas).

Cada indivíduo participante assinou um Termo de Consentimento Livre e Esclarecido (TCLE) (ANEXOS A e B) permitindo o uso de seu material biológico (amostra de sangue) para estudo. O material coletado foi armazenado em tubos rotulados de modo que o doador não pudesse ser diretamente identificado, mantendo, assim, o caráter confidencial da informação.

5.3 Determinação Fenotípica dos Indivíduos

A classificação dos indivíduos de acordo com a pigmentação da pele foi realizada segundo o sistema proposto por Fitzpatrick (Fitzpatrick 1988). Essa classificação se baseia nas características do indivíduo associadas à reação do mesmo à exposição ao sol, resultando na determinação do tipo de pele em seis grupos:

Tipos de pele	Descrição
I	Pele extremamente clara, sempre se queima, nunca bronzeia
II	Pele clara, sempre se queima, bronzeia às vezes
III	Pele média, algumas vezes se queima, sempre bronzeia
IV	Pele morena clara, raramente se queima, sempre bronzeia
V	Pele morena moderadamente pigmentada, nunca se queima, sempre bronzeia
VI	Pele negra marcadamente pigmentada, nunca se queima, sempre bronzeia

Adicionalmente, a intensidade da pigmentação da pele foi quantificada de forma objetiva por meio do espectrofotômetro portátil CM-2600d/2500d (*Konica Minolta Holdings, Inc*). A cor dos olhos foi caracterizada e dividida em categorias compreendendo azul, verde, mel, castanho-claro e castanho-escuro. A cor dos cabelos, por sua vez, foi caracterizada e dividida em categorias compreendendo ruivo, loiro-claro, loiro-escuro, castanho-claro, castanho-escuro e preto. A classificação da pele, cabelos e olhos foi sempre realizada por dois dos cinco pesquisadores do grupo, em comum acordo, durante as coletas das amostras, para diminuir a subjetividade do processo de heteroclassificação. Essas classificações eram feitas independentemente por cada um dos pesquisadores e quando não havia consenso imediato era realizada uma discussão entre a dupla para estabelecer um comum acordo. A classificação autodeclarada pelos indivíduos também foi levantada.

5.4 Análise Laboratorial

5.4.1 Extração de DNA

A extração do DNA das amostras de sangue foi realizada através de modificação do protocolo de *salting-out* (Miller, Dykes et al. 1988), como descrito no ANEXO D – item 1. De forma concisa, o processo consiste em etapas consecutivas de lise das hemácias, precipitação de proteínas (que serão descartadas) e lise dos linfócitos para expor o DNA.

5.4.2 Quantificação e verificação da qualidade do DNA genômico

Após extração do DNA, foi realizada a quantificação por eletroforese em gel de agarose, fluorescência (*Qubit – Life Technologies*) e medidas espectrofotométricas (*Thermo Scientific NanoDrop™ 1000 Spectrophotometer*).

A eletroforese em gel de agarose permitiu avaliar o nível de degradação do DNA genômico. O *NanoDrop®* foi utilizado para estimar a concentração inicial do DNA e avaliar o grau de contaminação do material genômico por proteínas. Já o *Qubit™* foi utilizado para avaliar a concentração de DNA genômico por meio do *dsDNA BR Assay (Invitrogen, Carlsbad, CA)*, de modo que diluições permitissem que este ficasse a uma concentração de 5ng/μL, normalizando assim as concentrações para a utilização deste na montagem das bibliotecas (ANEXO D – item 2).

5.4.3 Preparação das Bibliotecas

Após a quantificação, bibliotecas para cada amostra de DNA foram preparadas utilizando-se o *HaloPlex Target Enrichment System Kit*. Para a montagem de cada biblioteca são necessárias as seguintes etapas: digestão do DNA genômico por enzimas de restrição e hibridização dos fragmentos desejados com sondas biotinizadas cujas extremidades fossem complementares a esses fragmentos-alvo (durante esse passo os fragmentos são circularizados e índices, que irão identificar cada indivíduo durante o sequenciamento, são adicionados), captura desses fragmentos utilizando *beads* magnéticas recobertas por estreptavidina, ligação dos fragmentos circularizados, eluição e, finalmente, PCR dos fragmentos-alvo capturados. Após a amplificação, as bibliotecas de fragmentos geradas para cada indivíduo foram purificadas por meio de *beads* paramagnéticas utilizando-se o kit *Agencourt® AMPure® XP (Beckman Coulter, Inc)*. Todas estas etapas estão descritas detalhadamente no ANEXO D – itens 3, 4 e 5.

O controle de qualidade das bibliotecas e a normalização das mesmas foi realizado pelo sistema de eletroforese capilar *2100 Bioanalyzer (Agilent Technologies)* e por fluorescência (*Qubit™ dsDNA BR Assay, Life Technologies*), respectivamente (Anexo D – item 6).

5.4.4 Sequenciamento MiSeq

O sequenciamento foi realizado junto ao Departamento de Clínica Médica da FMRP-USP, por meio da plataforma *MiSeq Personal Sequencer (Illumina Inc.)*

utilizando o *MiSeq Reagent Kit V3 (600 cycle)*, seguindo-se as recomendações do fabricante, como descrito no ANEXO D – item 7. O sequenciamento foi do tipo leitura pareada das duas extremidades de cada fragmento (*paired-end*), aumentando o tamanho das sequências obtidas e melhorando a qualidade do mapeamento dessas sequências à sequência de referência.

5.5 Análises de Dados e Estatísticas

Como já mencionado anteriormente, após o sequenciamento, o fluxograma de análises de dados pode ser sumarizado em duas fases principais.

5.5.1 Pré-processamento e chamada de variantes

Esta fase tem início com o processamento dos dados brutos originados durante o sequenciamento. Uma pontuação de qualidade associada (*Q-score*) a cada base sequenciada é gerada e essas informações são armazenadas em um arquivo FastQ e disponibilizadas no ambiente de computação em nuvem da *Illumina, BaseSpace* (<https://basespace.illumina.com>).

No arquivo FastQ, o *Q-score* (também denominada *Phred-score*) se relaciona à probabilidade de erro na determinação de uma base. Por exemplo, um Q30 indica probabilidade de ocorrência de um erro a cada 1.000 bases determinadas. Maiores informações e detalhes sobre esta etapa estão descritos no ANEXO E – item 1.1.

A. AVALIAÇÃO DA QUALIDADE E TRIMAGEM

Posteriormente à obtenção e processamento dos dados brutos, é necessário verificar a qualidade das *reads* sequenciadas e realizar a trimagem (remoção) das *reads* de baixa qualidade para que o posterior mapeamento ocorra de maneira adequada.

O *software* FastQC (Andrews 2010) foi empregado para gerar um relatório completo da qualidade das sequências presentes em um arquivo FastQ. A trimagem foi realizada com uso da ferramenta *cutadapt* (Martin 2011) (maiores informações e detalhes estão presentes no ANEXO E – item 1.2).

B. ALINHAMENTO AO GENOMA

A etapa seguinte à conferência da qualidade e trimagem das *reads* é o mapeamento contra um genoma de referência (maiores informações e detalhes sobre esta etapa estão presentes no ANEXO E – item 1.2). Para isso, o genoma de referência (*Homo sapiens*, UCSC, hg19), disponibilizado pela *Illumina* foi indexado por meio do índice FM (Ferragina, Manzini et al. 2004).

Um pacote de softwares, denominado BWA, foi utilizado para realizar esse alinhamento. O BWA usa um algoritmo do tipo *Burrows-Wheeler* (BWT) que faz um processamento estatístico de reconstrução de um bloco de dados original de forma rápida e baixo consumo de memória (Li and Durbin 2009). O *output* originado é um arquivo SAM (*.sam) que foi convertido pelo programa *SAMtools* (<http://samtools.sourceforge.net/>) em um arquivo BAM (*.bam) porque este último é menor que o SAM e pode ser visualizado em programas como o IGV ou o *UCSC Genome Browser*. Neste trabalho utilizamos o *Integrative Genomics Viewer* (IGV) 2.3.66, um programa em java disponibilizado pelo *Broad Institute* (Robinson, Thorvaldsdottir et al. 2011, Thorvaldsdottir, Robinson et al. 2013). Um exemplo dessa visualização é apresentado na FIGURA 6.

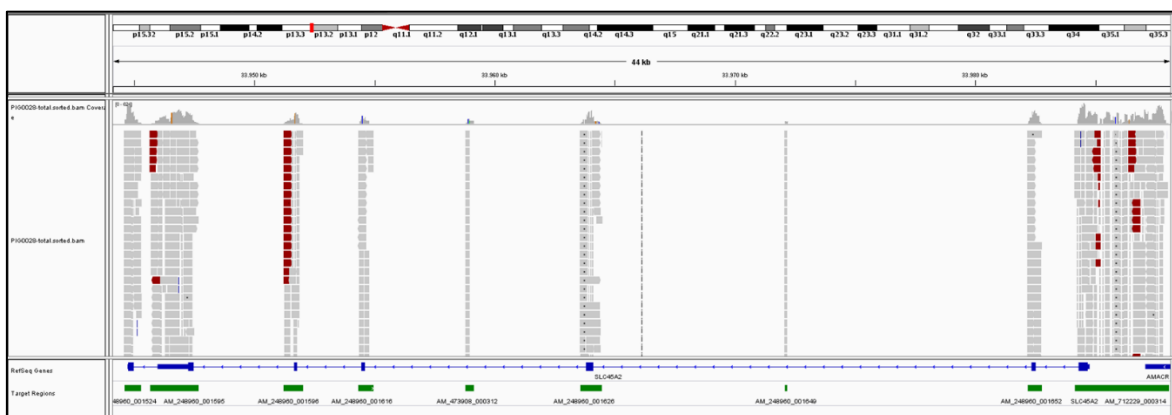


FIGURA 6. Representação dos alinhamentos das *reads* na região do gene *SLC45A2* no programa IGV_2.3.66. A parte principal da tela de visualização é dividida em 3 partes. No topo pode ser vista a representação esquemática do cromossomo 5 e a região deste cromossomo que está sendo visualizada (indicada por uma barra vermelha). Logo abaixo, é apresentada uma escala contendo as coordenadas da região ampliada. Na parte intermediária da tela é mostrada parte do arquivo BAM carregado. A visualização desse arquivo é dividida em duas regiões; acima, temos um gráfico representativo da cobertura (ou número de *reads* cobrindo cada base do genoma) e abaixo, os alinhamentos das *reads*. Na parte inferior da tela temos a anotação do genoma para a região visualizada (*RefSeq Genes*) em azul e em verde, a anotação das regiões cobertas pelo nosso ensaio.

Após o alinhamento das *reads* ao genoma de referência, o programa SAMStat (Lassmann, Hayashizaki et al. 2011) foi utilizado para verificar a acurácia do mapeamento. O programa utiliza o *mapping quality score* (MAPQ) para quantificar a probabilidade de uma *read* ter sido mapeada erroneamente no genoma.

C. DETERMINAÇÃO DE VARIANTES

O processo de determinação de variantes foi realizado por meio da comparação das regiões genômicas sequenciadas com um genoma de referência. O programa utilizado para tal análise foi o *Genome Analysis Toolkit* (GATK) v.3.5, disponibilizado pelo *Broad Institute* (McKenna, Hanna et al. 2010, DePristo, Banks et al. 2011, Van der Auwera, Carneiro et al. 2013) (maiores informações e detalhes sobre esta etapa estão presentes no ANEXO E – item 1.3).

Para realizar a determinação de SNPs foi utilizada a ferramenta *HaplotypeCaller* do GATK. Utilizando o *HaplotypeCaller*, foi gerado um arquivo GVCF (*.g.vcf) com o objetivo de obter informações adicionais sobre os sítios de variação. Esse arquivo compara todas as posições das *reads*, mesmo as não polimórficas com uma referência. Isso permite a obtenção de uma estimativa confiável em determinar se os sítios são homocigotos referência ou não (Van der Auwera, Carneiro et al. 2013).

D. REMOÇÃO DE GENÓTIPOS DE BAIXA QUALIDADE

Quando um segmento com baixa cobertura de sequenciamento ou quando um dos alelos é pouco representado (no caso de genótipos heterocigotos), a ferramenta *HaplotypeCaller* com o parâmetro `--emitRefConfidence GVCF` pode considerar incorretamente alguns genótipos.

Esses genótipos de baixa cobertura e alelos foram interrogados (indicado por um ponto, simbolizando uma incerteza), submetendo o arquivo VCF obtido pelo *HaplotypeCaller* (GVCF) à rotina do *script* VCFx (maiores informações e detalhes sobre esta etapa estão presentes no ANEXO E – item 1.3) (Castelli, Mendes-Junior et al. 2015, Lima, Buttura et al. 2016).

5.5.2 Inferência computacional de haplótipos

Para ambos os genes, o programa PHASE (Stephens, Smith et al. 2001) foi utilizado para a reconstrução dos haplótipos (ANEXO F – item 2). A abordagem de faseamento fragmentado foi utilizada para os genes *TYR* e *SLC45A2* como descrito no protocolo de análises estatísticas (ANEXO F – item 2.2) com algumas modificações. São elas: a probabilidade mínima adotada para aceitar como correto o faseamento para cada região foi estabelecida em 95% (ao invés de 80% como apresentado no protocolo) e a nomeação dos haplótipos não foi feita por meio de uma *network* e sim de acordo com a frequência dos mesmos.

Assim sendo, nessa abordagem, cada região gênica é inferida separadamente. Para o gene *SLC45A2* as variantes foram agrupadas em seis regiões, uma vez que

este gene possui três isoformas descritas (FIGURA 7): uma isoforma com o códon de parada no éxon sete (podendo ser observada uma 3'UTR menor); outra com um códon de parada no éxon seis (3'UTR maior); além da isoforma com um códon de parada no éxon cinco (esta isoforma tem apenas quatro éxons, pois além de não possuir os éxons seis e sete também não possui o éxon três). Dessa forma, as regiões analisadas foram: promotora (comum a todas as isoformas) com 24 pontos de variação; codificante com 7 éxons (isoforma 1) e 32 pontos de variação; codificante com 6 éxons (isoforma 2) e 20 pontos de variação; codificante com 4 éxons (isoforma 3) e 18 pontos de variação, 3'UTR maior (isoforma 2) com 9 pontos de variação e 3'UTR menor (isoforma 1) com dois pontos de variação. Não foi encontrado nenhum ponto de variação na região 5'UTR deste gene e na 3'UTR formada pela isoforma 3 com quatro éxons foi encontrado apenas um ponto de variação. Para o gene *TYR* as variantes foram agrupadas em duas regiões: região promotora com 20 pontos de variação e região codificante com 21 pontos de variação (nenhuma variante foi encontrada nas regiões 5'UTR e 3'UTR). Para ambos os genes (*TYR* e *SLC45A2*) foram considerados os éxons e suas regiões flanqueadoras cobertas por nosso ensaio (íntrons) como região codificante.

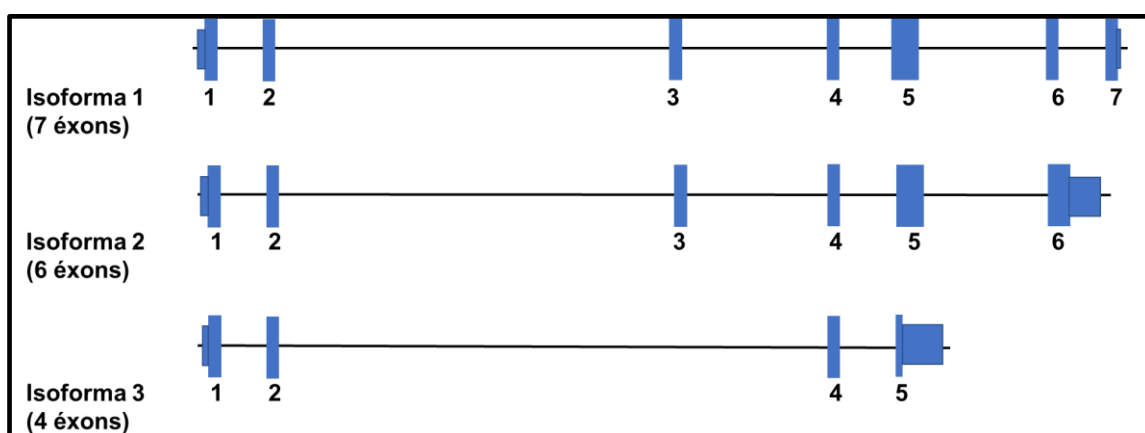


FIGURA 7. Representação esquemática das isoformas do gene *SLC45A2*. Os retângulos azuis representam os éxons e as regiões promotora (à esquerda) e 3'UTR (à direita). A isoforma 1 possui a 3'UTR de menor tamanho (31pb). A isoforma 2 possui 15pb a mais no éxon 6 e a 3'UTR de maior tamanho (1280pb). A isoforma 3 não possui os éxons 3, 6 e 7; o éxon 5 tem apenas 25pb e sua 3'UTR tem 245pb (incluindo parte do éxon 5 das outras duas isoformas). A figura não está em escala.

A nomenclatura dos haplótipos para ambos os genes na região promotora utilizou o prefixo “promo”. Já para o gene *TYR*, foi utilizado o prefixo “cds” para região codificante com íntrons, enquanto que para o gene *SLC45A2* temos: “Iso1cds” para a codificante isoforma 1, “Iso2cds” para a codificante isoforma 2, “Iso3cds” para a codificante isoforma 3, “Iso2utr” para a 3'UTR da isoforma 2 e “Iso1utr” para a 3'UTR da isoforma 1.

5.5.3 Análises estatísticas

O teste exato de Fisher bi-caudal, implementado em planilha desenvolvida para esta finalidade (Mendes Junior, C. T.), foi utilizado para averiguar a existência de associações entre alelos, genótipos ou haplótipos e um determinado fenótipo. Para determinar a intensidade desta associação foi realizado o cálculo de *Odds Ratio* (OR) e seu intervalo de confiança (IC) de 95% empregando-se a mesma planilha. O valor OR permite quantificar a magnitude da associação entre um alelo/genótipo/haplótipo e determinado fenótipo por meio da razão entre a probabilidade de um indivíduo portador de um determinado alelo/genótipo/haplótipo apresentar uma determinada característica e a probabilidade de um indivíduo portador do mesmo alelo/genótipo/haplótipo não apresentar tal característica. Foram consideradas significativas as associações com valores de p menores ou iguais a 0,05; OR maiores ou iguais a 10 (com algumas exceções que serão discutidas posteriormente) e limite inferior do IC de 95% maior do que um.

O desequilíbrio de ligação (DL) foi verificado em ambos os genes (*TYR* e *SLC45A2*) pelo programa Haploview (Barrett, Fry et al. 2005), levando-se em consideração apenas os SNPs. As frequências alélicas, heterozigose observada e esperada, assim como a aderência ao equilíbrio de Hardy-Weinberg foram calculadas pelo programa Arlequin (Excoffier e Lischer 2010) para as 200 amostras coletadas aleatoriamente. Estes cálculos foram realizados *locus-a-locus* para ambos os genes.

Mais informações sobre esta etapa estão presentes no ANEXO F – itens 3, 4 e 5.

5.5.4 Análise dos índices de melanina

Considerando apenas os SNPs que tenham se mostrado associados a alguma das características fenotípicas pela estratégia apresentada no item anterior, a existência de associação entre genótipos de SNPs dos genes *TYR* e *SLC45A2* e a variação da média no índice de melanina em cada genótipo foi testada por meio dos testes de Kruskal-Wallis e Mann-Whitney, utilizando-se o programa GraphPad InStat versão 3.06, GraphPad Software (www.graphpad.com). Apenas os dados dos genótipos da região da testa de dois SNPs (rs35407 e rs183671) do gene *SLC45A2* passaram no teste de normalidade; dessa forma, para manter a uniformidade das análises, foram realizados os testes não paramétricos para todos os SNPs, incluindo

os dois citados. Essa análise foi realizada em 335 das 340 amostras coletadas, pois informações sobre o índice de melanina não foram coletadas para cinco das amostras.

A distribuição dos índices de melanina, tanto para região exposta ao sol (testa) quanto para região pouco exposta ao sol (região interna do braço), também foi comparada entre os grupos de pele clara (I+II), pele média (III+IV) e pele escura (V+VI). O teste de Kruskal-Wallis (pKW) foi aplicado para comparar a variação das medianas quando havia informações dos três genótipos possíveis (homozigoto referência, heterozigoto e homozigoto alternativo), além dos três grupos de pele e o teste de Mann-Whitney quando havia informação de apenas dois genótipos diferentes.

O teste de Mann-Whitney avalia se uma variável tende a ter valores maiores em uma de duas amostras populacionais independentes. Ele é frequentemente usado como uma alternativa ao teste t quando os dados não estão normalmente distribuídos. Porém, enquanto o teste t usa as médias de uma amostra populacional, o teste de Mann-Whitney usa as medianas para sua análise (Hart 2001). Já o teste de Kruskal-Wallis é uma extensão do teste de Mann-Whitney para três ou mais grupos. Da mesma forma, ele compara medianas ao invés de médias, porém nele a hipótese nula é que as medianas das amostras populacionais são iguais e a hipótese alternativa é que existe diferença entre pelo menos duas delas (Bewick, Cheek et al. 2004).

5.5.5 Análise de miRNAs

O software mirhunt (Porto, Mendes-Junior et al. 2015) foi utilizado para procurar por sítios de ligação de miRNAs nas regiões 3'UTR do gene *SLC45A2*. MicroRNAs podem regular a expressão gênica a partir da interação com os diferentes haplótipos encontrados nessas regiões. O software mencionado realiza uma análise automatizada para procurar por interações entre a sequência alvo e miRNAs por meio de três algoritmos diferentes: RNAhybrid (Rehmsmeier, Steffen et al. 2004), miRanda (John, Enright et al. 2004) e intaRNA (Busch, Richter et al. 2008) e gera um relatório com a combinação dos resultados encontrados. As interações encontradas são ordenadas segundo sua especificidade (S), estabilidade ou energia de ligação (E) e uma combinação dos valores anteriores (ES).

Essa análise foi realizada considerando as regiões 3'UTR de cada uma das três isoformas: sete haplótipos com frequência maior que 1% da região 3'UTR da isoforma 2, isto é, a isoforma de maior tamanho, três haplótipos da região 3'UTR da região menor (incluindo dois haplótipos raros com frequência menor do que 1%) e

para os dois haplótipos (formados pelos dois alelos do único SNP - rs16891982) encontrados na região 3'UTR da isoforma 3.

Maiores informações sobre esta etapa estão presentes no ANEXO G.

6. RESULTADOS

6.1 Caracterização populacional e sequenciamento

A população de estudo constituiu um total de 340 indivíduos. Estes foram separados entre homens e mulheres e também estratificados de acordo com a forma de recrutamento para coleta: amostras coletadas aleatoriamente (204 indivíduos, quatro deles com alguma relação de parentesco com outros na nossa amostra) e amostras coletadas visando aumentar a amostragem de fenótipos de interesse para a pesquisa que estavam pouco representados na amostra aleatória (136 indivíduos). Em vista disso, a classificação amostral constituiu-se de: 167 mulheres e 173 homens; 144 indivíduos de pele clara (I e II na escala Fitzpatrick), 140 indivíduos de pele moderadamente pigmentada (III e IV na escala Fitzpatrick) e 56 indivíduos de pele escura (V e VI na escala Fitzpatrick). Quanto à pigmentação dos olhos, 37 indivíduos possuem olhos azuis, 62 possuem olhos verdes, 19 possuem olhos cor de mel, 73 possuem olhos castanho-claros e 149 possuem olhos castanho-escuros. Quanto à pigmentação dos cabelos, 11 indivíduos apresentam cabelos ruivos, 13 apresentam cabelos loiro-claros e 36 cabelos loiro-escuros, 61 indivíduos possuem cabelos castanho-claros, 125 cabelos castanho-escuros e 86 indivíduos possuem cabelos pretos; 8 indivíduos não tiveram a cor dos cabelos determinada. Quanto à presença ou ausência de sardas, 257 indivíduos possuíam algum tipo de mancha por exposição ao sol, pintas ou sardas e 83 indivíduos não possuíam. De acordo com a classificação baseada nos critérios adotados pelo IBGE (Instituto Brasileiro de Geografia e Estatística), 221 indivíduos foram classificados pelo grupo de pesquisa como brancos, 64 pardos, 44 pretos, 11 como amarelos e nenhum como indígena, enquanto que 218 doadores se auto classificaram como brancos, 67 como pardos, 42 como pretos, 12 como amarelos e 1 como indígena.

As características avaliadas na amostra populacional (estratificada em amostra aleatória e amostra direcionada para aumentar os fenótipos de interesse), isto é, cores de olhos, cabelos e pele e presença ou ausência de sardas estão representadas na Tabela 1.

Tabela 1: Características da amostra populacional de acordo com a heteroclassificação.

Características	Amostras aleatórias		Amostras visando aumentar a amostragem de fenótipos de interesse	
	Mulheres 104	Homens 100	Mulheres 63	Homens 73
Cor dos olhos				
Azul	4	9	12	12
Verde	12	10	18	22
Mel	5	3	6	5
Castanho-claro	28	34	4	7
Castanho-escuro	55	44	23	27
Cor dos cabelos*				
Ruivo	0	0	5	6
Loiro	11	13	17	8
Castanho	79	50	29	28
Preto	13	36	7	30
Cor da pele				
Clara (I + II)	38	34	37	35
Intermediária (III + IV)	53	54	18	15
Escura (V + VI)	13	12	8	23
Sardas				
Presença	84	69	57	47
Ausência	20	31	6	26

*Mulheres que tingem o cabelo e não estavam com a raiz exposta, assim como homens completamente calvos ou com a cabeça raspada ficaram sem classificação nesta categoria, somando um total de oito indivíduos (seis mulheres e dois homens).

O sequenciamento foi realizado para as 340 amostras em cinco ensaios. As amostras que não apresentaram cobertura satisfatória nos primeiros ensaios foram sequenciadas novamente em ensaios posteriores. Podemos observar na Tabela 2 que todos os sequenciamentos foram bem sucedidos, com valores de %PF (*clusters* que passaram nos filtros de qualidade) acima de 75%, alta quantidade de *reads* por sequenciamento e que a porcentagem de bases sequenciadas com qualidade *Phred* 30 ou superior (%Q30 Médio) ficou em torno ou acima de 55%. É importante ressaltar que esta última coluna não reflete adequadamente a qualidade do sequenciamento, já que *amplicons* ainda que bem menores que 300pb acabam resultando em *reads* compostas por 301 nucleotídeos (uma vez que a cada ciclo ocorre a leitura de um nucleotídeo teoricamente incorporado, ainda que na prática isso não ocorra), prejudicando a qualidade geral das *reads*. Após processadas pelo *cutadapt* (Martin 2011), uma maior proporção das *reads* compostas por nucleotídeos de elevada qualidade foi obtida (ANEXO E – item 1.2).

Tabela 2: Dados do aproveitamento e qualidade dos cinco ensaios de sequenciamento.

	%PF	Nº de <i>reads</i>	%Q30 Médio
Sequenciamento 1 (<i>n</i> =48)	78,79	48,4 milhões	55,65
Sequenciamento 2 (<i>n</i> =95)	82,08	47,6 milhões	57,80
Sequenciamento 3 (<i>n</i> =87)	96,72	21,3 milhões	62,67
Sequenciamento 4 (<i>n</i> =80)	84,07	45,1 milhões	54,96
Sequenciamento 5 (<i>n</i> =60)	85,74	43,6 milhões	56,28

6.2 Chamada de variantes e haplótipos inferidos

A ferramenta *HaplotypeCaller*, utilizada com o parâmetro `--emitRefConfidence GVCF` (GATK), encontrou 104 pontos de variação para o gene *SLC45A2* e 81 para o gene *TYR*. A rotina do VCFx foi utilizada nesses arquivos e os pontos de variação caracterizados por baixa cobertura foram interrogados. Após a análise dos relatórios e arquivos produzidos pelo VCFx e uma verificação ponto a ponto das variantes, foram removidos das análises posteriores os *singletons* (pontos de variação nos quais apenas uma cópia do alelo alternativo estava presente) e as regiões que apresentaram baixa cobertura e *reads* com baixa qualidade. Assim, para o gene *SLC45A2* foram removidos 39 *singletons* e sete regiões, restando 58 pontos de variação para as análises posteriores. Para o gene *TYR* foram retirados 25 *singletons* e 14 regiões de baixa qualidade, permanecendo 42 sítios.

Como já mencionado, para ambos os genes foram considerados corretos apenas os haplótipos faseados com probabilidade igual ou superior a 95%, o que garante a utilização subsequente apenas das amostras cujos haplótipos tenham sido definidos com grande suporte estatístico. Assim sendo, de acordo com as seis regiões analisadas do gene *SLC45A2*, foram encontrados: 21 haplótipos em 331 indivíduos na região promotora (24 pontos de variação) (Tabela 3); 36 haplótipos em 307 indivíduos na região codificante da isoforma 1 (32 pontos de variação) (Tabela 4); 28 haplótipos em 307 indivíduos na região codificante da isoforma 2 (20 pontos de variação) (Tabela 5); 24 haplótipos em 310 indivíduos na região codificante da isoforma 3 (18 pontos de variação) (Tabela 6); 10 haplótipos em 337 indivíduos na região 3'UTR da isoforma 2 (nove pontos de variação) (Tabela 7) e 3 haplótipos em 340 indivíduos na região 3'UTR da isoforma 1 (dois pontos de variação) (Tabela 8). Para o gene *TYR* foram encontrados 19 haplótipos diferentes na região promotora (20 pontos de variação) em 305 indivíduos (Tabela 9) e 19 haplótipos diferentes na região codificante (21 pontos de variação) em 327 indivíduos (Tabela 10).

Tabela 3: Haplótipos encontrados para região promotora do gene *SLC45A2* considerando 24 pontos de variação em 331 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipos	rs548644095	33985178	rs369350969	rs112118977	rs182613292	rs551953158	rs6867641	rs67752819	rs191460200	rs184103040	rs13289	rs148899183	rs137931037	rs73077153	rs115214355	rs73077154	rs143386685	rs182361997	rs529140346	rs186629199	rs113746028	rs563065271	rs116206502	rs116333833	Frequência	
promo1	TTCTC	A	TTC	T	T	C	C	C	G	G	G	G	C	C	T	G	G	T	G	C	C	G	C	C	0,4698	
promo2	T	CATT	.	.	C	0,3172
promo3	CATT	.	.	C	0,1057
promo4	CATT	.	.	C	C	0,0196
promo5	CATT	.	.	C	A	.	0,0196
promo6	.	.	.	A	.	.	.	CATT	.	.	C	T	.	.	.	0,0076
promo7	CATT	.	.	C	T	.	.	.	0,0045
promo8	C	A	0,0060
promo9	A	0,0060
promo10	CATT	.	.	C	C	T	.	.	.	0,0030
promo11	CATT	.	.	C	A	.	.	.	C	0,0045
promo12	CATT	.	.	C	.	T	G	0,0045
promo13	CATT	T	.	C	0,0045
promo14	A	0,0045
promo15	T	CATT	.	.	C	A	0,0045
promo16	T	CATT	0,0030
promo17	CATT	.	.	C	.	.	T	.	C	0,0030
promo18	T	0,0030
promo19	C	.	.	CATT	.	.	C	A	.	.	.	C	0,0030
promo20	T	CATT	.	.	C	T	0,0030
promo21	.	.	T	.	.	.	T	CATT	.	.	C	0,0030

Tabela 4: Haplótipos encontrados para região codificante da isoforma 1 do gene *SLC45A2* considerando 32 pontos de variação em 307 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipos	rs535871299	rs146922924	rs143719822	rs189917057	rs35406	rs7380538	rs35407	rs57959046	rs55746663	rs114794536	rs6874998	rs6894408	rs250416	rs16891982	rs35398	rs560973795	rs116828782	rs113928135	rs2287949	rs35990319	rs147962674	rs1010872	rs28777	rs77366462	rs141348470	rs26722	rs376784068	rs183671	rs6451048	rs6451049	rs58784954	rs250413	Frequências	
Iso1cds1	C	C	T	A	A	G	G	C	A	T	T	C	C	G	G	TAGCCCA	G	G	C	G	GTGTATTATT	C	A	G	A	C	T	G	G	C	G	G	0,5945	
Iso1cds2	G	.	A	A	C	T	T	T	0,0733	
Iso1cds3	A	C	T	0,0570	
Iso1cds4	C	0,0521	
Iso1cds5	A	A	C	T	.	.	A	C	.	AAAC	.	.	.	T	.	T	.	0,0554	
Iso1cds6	A	A	G	C	T	0,0309	
Iso1cds7	G	.	A	A	C	T	T	.	T	.	.	.	A	0,0098	
Iso1cds8	A	0,0195
Iso1cds9	A	0,0065	
Iso1cds10	.	CA	A	A	.	.	C	G	.	C	A	C	.	.	T	.	T	0,0098	
Iso1cds11	G	.	A	A	C	T	T	0,0081	
Iso1cds12	G	.	A	A	C	T	T	T	.	T	0,0081	
Iso1cds13	A	A	C	T	.	.	A	C	.	AAAC	.	.	T	A	T	.	.	0,0065	
Iso1cds14	T	A	A	C	T	.	.	A	C	.	AAAC	.	.	T	.	T	.	.	0,0065	
Iso1cds15	A	C	.	.	T	T	0,0049	
Iso1cds16	G	.	C	.	.	.	A	T	0,0049	
Iso1cds17	.	CA	A	G	.	C	T	0,0049	
Iso1cds19	.	CA	C	T	0,0033	
Iso1cds20	A	A	C	G	T	0,0033	
Iso1cds22	A	A	G	.	C	A	C	.	.	.	T	0,0033	
Iso1cds23	.	CA	C	G	.	C	A	.	.	T	0,0033	
Iso1cds24	AAAC	0,0033	
Iso1cds25	.	CA	A	A	G	C	T	0,0033	
Iso1cds26	T	.	T	.	.	.	0,0033	
Iso1cds27	.	CA	A	A	C	T	.	T	0,0033	
Iso1cds28	C	T	0,0033	
Iso1cds29	A	C	T	.	T	0,0033	
Iso1cds30	.	.	.	T	0,0016	
Iso1cds31	C	A	0,0016	
Iso1cds36	A	A	C	T	.	.	A	C	.	.	.	T	0,0016	
Iso1cds37	A	A	C	T	.	.	A	C	0,0016	
Iso1cds38	G	.	A	A	C	T	T	.	T	.	.	A	0,0016	
Iso1cds39	G	.	A	A	C	T	T	.	T	.	.	.	0,0016	
Iso1cds40	AAAC	.	.	T	.	T	.	.	.	0,0016

Tabela 5: Haplótipos encontrados para região codificante da isoforma 2 do gene *SLC45A2* considerando 20 pontos de variação em 307 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipos	rs250416	rs16891982	rs35398	rs560973795	rs116828782	rs113928135	rs2287949	rs35990319	rs147962674	rs1010872	rs28777	rs77366462	rs141348470	rs26722	rs376784068	rs183671	rs6451048	rs6451049	rs58784954	rs250413	Frequências
Iso2cds1	C	G	G	TAGCCCA	G	G	C	G	GTTGTATTTATTT	C	A	G	A	C	T	G	G	C	G	G	0,6189
Iso2cds2	.	C	C	.	.	.	T	0,1059
Iso2cds3	A	C	T	C	.	.	T	0,0733
Iso2cds4	.	C	T	.	.	A	C	.	AAAC	.	.	T	.	T	.	.	0,0619
Iso2cds5	.	C	0,0554
Iso2cds6	A	C	T	C	.	.	T	.	T	.	.	.	A	0,0049
Iso2cds7	.	C	A	C	.	.	T	.	T	0,0081
Iso2cds8	A	0,0065
Iso2cds9	A	C	T	C	.	.	.	T	0,0049
Iso2cds10	A	C	T	T	C	.	.	T	.	T	0,0081
Iso2cds11	.	C	.	.	.	A	C	.	.	.	T	0,0081
Iso2cds12	.	C	T	.	.	A	C	.	AAAC	.	.	T	A	T	.	.	0,0065
Iso2cds13	.	C	.	.	T	C	.	.	.	T	0,0049
Iso2cds15	.	C	G	.	C	A	.	.	T	0,0016
Iso2cds16	.	C	C	.	.	T	.	T	0,0033
Iso2cds18	AAAC	.	.	T	0,0033
Iso2cds19	.	C	G	.	C	.	.	.	T	0,0016
Iso2cds21	T	.	T	0,0033
Iso2cds23	A	C	C	.	.	T	.	T	0,0033
Iso2cds24	.	C	T	0,0033
Iso2cds25	.	C	A	0,0016
Iso2cds27	.	C	T	.	.	A	C	.	.	.	T	0,0016
Iso2cds29	.	C	C	0,0016
Iso2cds30	.	C	T	.	.	A	C	0,0016
Iso2cds31	.	C	A	.	0,0016
Iso2cds32	AAAC	.	T	.	T	.	.	.	0,0016
Iso2cds33	A	C	T	C	.	AAAC	.	T	.	T	.	A	.	0,0016
Iso2cds35	A	C	T	C	.	AAAC	.	T	.	T	.	.	.	0,0016

Tabela 6: Haplótipos encontrados para região codificante da isoforma 3 do gene *SLC45A2* considerando 18 pontos de variação em 310 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipos	rs35398	rs560973795	rs116828782	rs113928135	rs2287949	rs35990319	rs147962674	rs1010872	rs28777	rs77366462	rs141348470	rs26722	rs376784068	rs183671	rs6451048	rs6451049	rs58784954	rs250413	Frequências
Iso3cds1	G	TAGCCCA	G	G	C	G	GTTGTATTATT	C	A	G	A	C	T	G	G	C	G	G	0,6742
Iso3cds2	C	T	0,1065
Iso3cds3	T	C	.	.	T	.	T	0,0710
Iso3cds4	T	.	.	A	C	.	AAAC	.	.	T	.	T	.	.	0,0613
Iso3cds5	T	C	.	.	T	.	T	.	.	.	A	0,0048
Iso3cds6	A	0,0097
Iso3cds7	A	C	.	.	T	.	T	0,0081
Iso3cds8	T	C	T	0,0048
Iso3cds9	T	T	C	.	.	T	.	T	0,0081
Iso3cds10	AAAC	0,0065
Iso3cds11	.	.	.	A	C	T	0,0081
Iso3cds12	C	.	.	T	.	T	0,0065
Iso3cds13	T	.	.	A	C	.	AAAC	.	.	T	A	T	.	.	0,0065
Iso3cds14	.	.	T	C	T	0,0048
Iso3cds16	G	.	C	T	0,0032
Iso3cds17	G	.	C	A	.	.	.	T	0,0016
Iso3cds18	A	.	0,0016
Iso3cds19	T	.	T	0,0032
Iso3cds20	T	0,0016
Iso3cds26	T	.	.	A	C	0,0016
Iso3cds27	C	0,0016
Iso3cds28	AAAC	.	.	T	.	T	.	.	0,0016
Iso3cds29	T	C	.	AAAC	.	.	T	.	T	.	A	0,0016
Iso3cds30	T	C	.	AAAC	.	.	T	.	T	.	.	0,0016

Tabela 7: Haplótipos encontrados para região 3'UTR da isoforma 2 do gene *SLC45A2* considerando 9 pontos de variação em 337 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipos	rs189917057	rs35406	rs7380538	rs35407	rs57959046	rs55746663	rs114794536	rs6874998	rs6894408	Frequências
Iso2utr1	A	A	G	G	C	A	T	T	C	0,6588
Iso2utr2	.	G	.	A	0,1098
Iso2utr3	.	.	.	A	0,0801
Iso2utr4	.	.	A	A	0,0712
Iso2utr5	.	.	.	A	A	G	.	.	.	0,0371
Iso2utr6	A	.	.	0,0193
Iso2utr7	.	.	A	A	.	.	.	C	G	0,0104
Iso2utr8	G	0,0074
Iso2utr9	.	.	A	A	G	0,0030
Iso2utr10	T	0,0030

Tabela 8: Haplótipos encontrados para região 3'UTR da isoforma 1 do gene *SLC45A2* considerando dois pontos de variação em 340 amostras.

Haplótipos	rs45552240	rs3822467	Frequências
Iso1utr1	T	C	0,9897
Iso1utr2	G	.	0,0059
Iso1utr3	.	T	0,0044

Tabela 9: Haplótipos encontrados para região promotora do gene *TYR* considerando 20 pontos de variação em 305 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipos	rs7924925	rs151276292	rs75741664	rs181833244	88907815	rs7925404	rs142881018	rs139927890	rs61759511	rs61759512	rs11824466	rs114470821	rs61759516	rs1160550168	rs563971364	rs143441124	rs5021654	rs4547091	rs55870815	rs1799989	Frequências
promo1	A	A	C	A	A	G	CCTGA	A	T	G	C	A	G	T	TA	C	G	C	T	C	0,5033
promo2	A	C	T	.	.	0,2295
promo3	G	0,0852
promo4	C	T	.	A	0,0557
promo5	A	C	T	.	A	0,0279
promo6	G	C	T	.	A	0,0033
promo7	G	C	T	.	.	0,0115
promo8	T	C	T	.	.	0,0164
promo9	.	.	T	AAT	.	A	C	T	.	.	0,0164
promo10	G	C	C	T	.	.	0,0115
promo11	C	T	.	.	0,0066
promo12	A	0,0066
promo13	G	C	C	T	.	.	0,0066
promo14	T	G	C	T	.	A	0,0049
promo15	G	C	C	T	C	.	0,0049
promo16	G	T	.	C	T	.	.	0,0033
promo17	A	0,0016
promo18	.	.	.	G	.	A	C	T	.	A	0,0033
promo19	T	0,0016

Tabela 10: Haplótipos encontrados para região codificante do gene *TYR* considerando 21 pontos de variação em 327 amostras. Variantes e haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipos	rs1939261	rs1939260	rs1042602	rs4987234	rs199569456	rs142575793	rs138538705	rs11018530	rs141481082	rs67512950	rs12804012	rs75375648	rs113438665	rs147210895	rs3793975	rs184834824	rs1393350	rs1126809	rs145841599	rs561489771	rs13312744	Frequências
cds1	G	C	A	T	A	T	T	C	T	ACTT	G	T	T	C	G	C	G	G	G	C	T	0,3058
cds2	.	.	C	0,2997
cds3	.	.	C	A	A	.	.	.	0,1713
cds4	.	.	C	G	.	A	A	.	.	.	A	0,1177
cds5	.	.	C	G	.	.	A	0,0153
cds6	.	.	C	.	.	.	G	0,0138
cds7	.	.	C	G	.	A	A	.	C	0,0122
cds8	A	T	C	A	0,0122
cds9	.	.	C	C	0,0122
cds10	A	.	.	.	0,0107
cds11	.	.	C	C	.	.	.	G	C	A	A	.	.	.	A	0,0076
cds12	.	.	C	A	0,0046
cds13	.	.	C	.	.	.	G	G	.	0,0046
cds14	A	A	.	.	.	0,0015
cds15	.	.	C	.	.	C	.	G	.	A	A	.	C	0,0031
cds16	C	0,0031
cds19	.	.	C	A	0,0015
cds20	.	.	C	G	.	A	A	.	.	.	A	.	A	A	.	.	.	0,0015
cds23	.	.	C	A	0,0015

6.3 Análises populacionais e desequilíbrio de ligação

Considerando somente os 200 indivíduos coletados aleatoriamente e a definição de *locus* polimórfico como sendo o *locus* que possui o alelo menos frequente com frequência maior que 1%, encontramos 25 variantes (das 58 analisadas) polimórficas para o gene *SLC45A2* e 19 (das 42 analisadas) para o gene *TYR*. Dessa forma, podemos observar nas Tabelas 11 e 12 que o gene *SLC45A2* possui 33 polimorfismos com o alelo menos frequente com frequência entre 0 e 1% (nove deles - rs143386685, rs73077153, rs551953158, rs58784954, rs6451048, rs376784068, rs113928135, rs6874998 e rs45552240 - associados com algum fenótipo avaliado) e o gene *TYR* possui 23 (quatro deles - 88907815, rs142881018, rs55870815, rs75375648 - associados com algum fenótipo avaliado).

Desconsiderando-se os SNPs não polimórficos, o menor valor de heterozigose observada (*Ho*) para o gene *SLC45A2* foi de 0,0400 no *locus* rs114794536, enquanto que o maior valor foi de 0,4472 no *locus* rs67752819. Para o gene *TYR* o menor valor foi de 0,0200 no *locus* rs11824466, enquanto o maior valor foi de 0,5202 no *locus* rs28521275. Já a heterozigose esperada (*He*) teve o menor valor (0,0393) para o mesmo *locus* de menor valor da *Ho* (rs114794536) para o gene *SLC45A2* e o maior valor (0,4997) para o *locus* rs13289. Para o gene *TYR*, o menor valor foi de 0,0248 em dois polimorfismos (rs75741664 e rs139927890) e o maior valor foi de 0,5010 no mesmo *locus* de maior valor da *Ho* (rs28521275).

Em relação ao equilíbrio de Hardy-Weinberg, observou-se que apresentaram desvios as variantes rs35407, rs16891982, rs28777 e rs183671 do gene *SLC45A2* e a variante rs11824466 do gene *TYR*.

Tabela 11: Frequências alélicas, heterozigose observada (*H_o*), heterozigose esperada (*H_e*) e aderência ao Equilíbrio de Hardy-Weinberg (*p_{EHW}*) para cada *locus* do gene *SLC45A2*. Valores significativos ($\alpha = 0,05$) na análise de aderência ao equilíbrio estão marcados em vermelho.

Posição	Região	Locus	Frequência		<i>H_o</i>	<i>H_e</i>	<i>p_{HW}</i>
			Alelo Referência	Alelo Alternativo			
33987471	promotora	rs116333833 (C/A)	0,9950	0,0050	0,0100	0,0100	1,0000
33987394	promotora	rs116206502 (C/A)	0,9700	0,0300	0,0500	0,0584	0,1561
33987346	promotora	rs563065271 (G/A)	0,9950	0,0025	0,0050	0,0050	1,0000
33987237	promotora	rs113746028 (C/T)	0,9750	0,0200	0,0402	0,0395	1,0000
33987219	promotora	rs186629199 (C/G)	0,9925	0,0025	0,0050	0,0050	1,0000
33987154	promotora	rs529140346 (G/T)	0,9950	0,0025	0,0050	0,0050	1,0000
33987110	promotora	rs182361997 (T/A)	0,9925	0,0050	0,0101	0,0100	1,0000
33986992	promotora	rs143386685 (G/A)	0,9950	0,0050	0,0100	0,0100	1,0000
33986871	promotora	rs73077154 (G/C)	0,9725	0,0275	0,0550	0,0536	1,0000
33986606	promotora	rs115214355 (T/C)	0,9950	0,0050	0,0100	0,0100	1,0000
33986605	promotora	rs73077153 (C/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33986567	promotora	rs137931037 (C/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33986410	promotora	rs148899183 (G/A)	0,9975	0,0025	0,0050	0,0050	1,0000
33986409	promotora	rs13289 (C/G)	0,5275	0,4725	0,4350	0,4997	0,0874
33986343	promotora	rs184103040 (G/A)	0,9925	0,0050	0,0101	0,0100	1,0000
33986123	promotora	rs191460200 (G/T)	0,9925	0,0075	0,0150	0,0149	1,0000
33985861	promotora	rs67752819 (CATT/C)	0,5325	0,4650	0,4472	0,4988	0,1585
33985857	promotora	rs6867641 (T/C)	0,3425	0,6550	0,4271	0,4526	0,4371
33985820	promotora	rs551953158 (C/T)	1,0000	0,0000	-	-	-
33985792	promotora	rs182613292 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
33985753	promotora	rs112118977 (T/A)	0,9900	0,0050	0,0101	0,0101	1,0000
33985726	promotora	rs369350969 (TTC/T)	0,9950	0,0050	0,0100	0,0100	1,0000
33985178	promotora	33985178 (A/G)	1,0000	0,0000	-	-	-
33984981	promotora	rs548644095 (TTCTC/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33982568	intron 1	rs250413 (G/A)	0,9625	0,0350	0,0603	0,0681	0,2118
33972116	intron 2	rs58784954 (G/A)	0,9925	0,0075	0,0150	0,0149	1,0000
33964352	intron 2	rs6451049 (T/C)	0,0650	0,9250	0,1320	0,1236	1,0000
33964241	intron 2	rs6451048 (A/G)	0,0025	0,9925	0,0051	0,0051	1,0000
33964210	intron 2	rs183671 (T/G)	0,3325	0,6625	0,3737	0,4456	0,0261

33964140	intron 2	rs376784068 (T/A)	0,9925	0,0025	0,0051	0,0051	1,0000
33963870	exon 3 (iso1 e 2)/intron 2 (iso3)	rs26722 (C/T)	0,8800	0,1200	0,2200	0,2117	0,7460
33963731	intron 3 (iso1 e 2)/intron 2 (iso3)	rs141348470 (AAAC/A)	0,0700	0,9275	0,1407	0,1311	0,6047
33958993	intron 3 (iso1 e 2)/intron 2 (iso3)	rs77366462 (G/A)	0,9950	0,0025	0,0050	0,0050	1,0000
33958959	intron 3 (iso1 e 2)/intron 2 (iso3)	rs28777 (C/A)	0,3250	0,6725	0,3568	0,4392	0,0103
33958910	intron 3 (iso1 e 2)/intron 2 (iso3)	rs1010872 (A/C)	0,0750	0,9250	0,1500	0,1391	0,6071
33958873	intron 3 (iso1 e 2)/intron 2 (iso3)	rs147962674 (GTTGTATTTATTT/G)	0,9900	0,0100	0,0200	0,0199	1,0000
33954546	exon 4	rs35990319 (G/A)	1,0000	0,0000	-	-	-
33954511	exon 4	rs2287949 (T/C)	0,0650	0,9350	0,1300	0,1219	1,0000
33954457	intron 4	rs113928135 (G/A)	0,9900	0,0100	0,0200	0,0199	1,0000
33954408	intron 4	rs116828782 (G/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33954387	intron 4	rs560973795 (TAGCCCA/T)	0,9950	0,0050	0,0100	0,0100	1,0000
33951826	intron 4	rs35398 (G/T)	0,8950	0,1050	0,1900	0,1884	1,0000
33951693	exon 5 (iso1 e 2)/3'UTR (iso3)	rs16891982 (C/G)	0,3925	0,6075	0,4050	0,4781	0,0374
33947544	intron 5 (iso1 e 2)	rs250416 (C/A)	0,8850	0,1100	0,2020	0,1980	1,0000
33947168	intron 6 (iso1)/3'UTR (iso2)	rs6894408 (C/G)	0,9775	0,0225	0,0450	0,0441	1,0000
33946949	intron 6 (iso1)/3'UTR (iso2)	rs6874998 (T/C)	0,9900	0,0100	0,0200	0,0199	1,0000
33946946	intron 6 (iso1)/3'UTR (iso2)	rs114794536 (T/A)	0,9800	0,0200	0,0400	0,0393	1,0000
33946931	intron 6 (iso1)/3'UTR (iso2)	rs55746663 (A/G)	0,9675	0,0325	0,0650	0,0631	1,0000
33946876	intron 6 (iso1)/3'UTR (iso2)	rs57959046 (C/A)	0,9675	0,0325	0,0650	0,0631	1,0000
33946571	intron 6 (iso1)/3'UTR (iso2)	rs35407 (A/G)	0,2975	0,7025	0,3350	0,4190	0,0068
33946555	intron 6 (iso1)/3'UTR (iso2)	rs7380538 (A/G)	0,0825	0,9175	0,1550	0,1518	1,0000
33946143	intron 6 (iso1)/3'UTR (iso2)	rs35406 (A/G)	0,8950	0,1050	0,1900	0,1884	1,0000
33946117	intron 6 (iso1)/3'UTR (iso2)	rs189917057 (A/T)	0,9950	0,0050	0,0100	0,0100	1,0000
33945915	intron6 (iso1)	rs143719822 (T/A)	0,9925	0,0050	0,0101	0,0100	1,0000
33945902	intron6 (iso1)	rs146922924 (C/CA)	0,9700	0,0275	0,0553	0,0539	1,0000
33945869	intron6 (iso1)	rs535871299 (C/T)	0,9900	0,0050	0,0101	0,0101	1,0000
33944752	3'UTR (iso1)	rs3822467 (C/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33944729	3'UTR (iso1)	rs45552240 (T/G)	0,9975	0,0025	0,0050	0,0050	1,0000

Tabela 12: Frequências alélicas, heterozigose observada (H_o), heterozigose esperada (H_e) e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) para cada *locus* do gene *TYR*. Valores significativos ($\alpha = 0,05$) na análise de aderência ao equilíbrio estão marcados em vermelho.

Posição	Região	Locus	Frequência		H_o	H_e	p_{HW}
			Alelo Referência	Alelo Alternativo			
88907621	promotora	rs7924925 (G/A)	0,1525	0,8475	0,2650	0,2591	1,0000
88907737	promotora	rs151276292 (A/C)	0,9950	0,0050	0,0100	0,0100	1,0000
88907781	promotora	rs75741664 (C/T)	0,9875	0,0125	0,0250	0,0248	1,0000
88907811	promotora	rs181833244 (A/G)	0,9950	0,0050	0,0100	0,0100	1,0000
88907996	promotora	rs7925404 (G/A)	0,7350	0,2650	0,3800	0,3905	0,7166
88908275	promotora	rs142881018 (CCTGA/C)	0,9900	0,0100	0,0200	0,0199	1,0000
88908457	promotora	rs139927890 (A/AAT)	0,9875	0,0125	0,0250	0,0248	1,0000
88908667	promotora	rs61759511 (T/C)	0,9900	0,0100	0,0200	0,0199	1,0000
88908789	promotora	rs61759512 (G/A)	0,9825	0,0125	0,0253	0,0250	1,0000
88908905	promotora	rs11824466 (C/T)	0,9800	0,0200	0,0200	0,0393	0,0014
88909024	promotora	rs114470821 (A/G)	0,9975	0,0025	0,0050	0,0050	1,0000
88909316	promotora	rs61759516 (G/A)	0,9875	0,0100	0,0201	0,0200	1,0000
88909753	promotora	rs1160550168 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
88909835	promotora	rs563971364 (TA/T)	0,9975	0,0025	0,0050	0,0050	1,0000
88910428	promotora	rs143441124 (C/A)	0,9950	0,0025	0,0050	0,0050	1,0000
88910589	promotora	rs5021654 (G/C)	0,5800	0,4175	0,4472	0,4883	0,2454
88910821	promotora	rs4547091 (C/T)	0,5800	0,4175	0,4472	0,4883	0,2466
88910882	promotora	rs55870815 (T/C)	0,9950	0,0025	0,0050	0,0050	1,0000
88910923	promotora	rs1799989 (C/A)	0,8750	0,1250	0,2200	0,2193	1,0000
88911235	exon 1	rs1939261 (G/A)	0,9925	0,0075	0,0150	0,0149	1,0000
88911299	exon 1	rs1939260 (C/T)	0,9925	0,0075	0,0150	0,0149	1,0000
88911696	exon 1	rs1042602 (C/A)	0,6750	0,3250	0,4100	0,4399	0,3395
88924385	exon 2	rs4987234 (T/C)	0,9925	0,0075	0,0150	0,0149	1,0000
88924611	intron 2	rs199569456 (A/AT)	0,9975	0,0025	0,0050	0,0050	1,0000
88924613	intron 2	rs142575793 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
88933411	intron 2	rs138538705 (T/G)	0,9400	0,0200	0,0430	0,0422	1,0000
88933629	intron 2	rs11018530 (C/G)	0,8450	0,1525	0,2663	0,2602	1,0000
88933658	intron 2	rs141481082 (T/C)	0,9900	0,0075	0,0151	0,0150	1,0000
88933717	intron 2	rs67512950 (ACTT/A)	0,8600	0,1400	0,2500	0,2414	0,7733
88960790	intron 2	rs12804012 (G/A)	0,8325	0,1650	0,2915	0,2774	0,6104

88960816	intron 2	rs75375648 (T/C)	0,9900	0,0075	0,0151	0,0150	1,0000
88960914	intron 2	rs113438665 (T/C)	0,9850	0,0125	0,0251	0,0249	1,0000
88961165	intron 3	rs147210895 (C/T)	0,9800	0,0050	0,0103	0,0103	1,0000
88961188	intron 3	rs3793975 (G/A)	0,8600	0,1250	0,2268	0,2251	1,0000
89010998	intron 3	rs184834824 (C/T)	0,9950	0,0050	0,0100	0,0100	1,0000
89011046	intron 3	rs1393350 (G/A)	0,8200	0,1800	0,2700	0,2959	0,2320
89017961	exon 4	rs1126809 (G/A)	0,8025	0,1850	0,2806	0,3039	0,3433
89018281	intron 4	rs145841599 (G/A)	0,9575	0,0075	0,0160	0,0160	1,0000
89028204	intron 4	rs561489771 (C/G)	0,9950	0,0050	0,0100	0,0100	1,0000
89028323	exon 5	rs13312744 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
89028969	3' <i>Downstream</i>	rs28521275 (C/T)	0,4475	0,4775	0,5202	0,5010	0,6510

O desequilíbrio de ligação foi avaliado para os genótipos não faseados em ambos os genes. Para o gene *SLC45A2* foram considerados 51 SNPs nos 340 indivíduos. Podemos observar na FIGURA 8 dois blocos em desequilíbrio, o primeiro incluindo da porção final do íntron 2 à porção final da 3'UTR da isoforma 2, composto por 21 pontos de variação e o segundo, abrangendo parte da região promotora, composto por 4 pontos de variação. Já para o gene *TYR* foram considerados 37 SNPs nos 340 indivíduos. Podemos observar na FIGURA 9 dois blocos em desequilíbrio, o primeiro abrangendo parte final da região promotora até a porção inicial do íntron 3, composto por 17 pontos de variação e o segundo abrangendo toda a porção final do gene, desde o final do íntron 3 até o SNP 3'*Downstream*, composto por 6 pontos de variação.

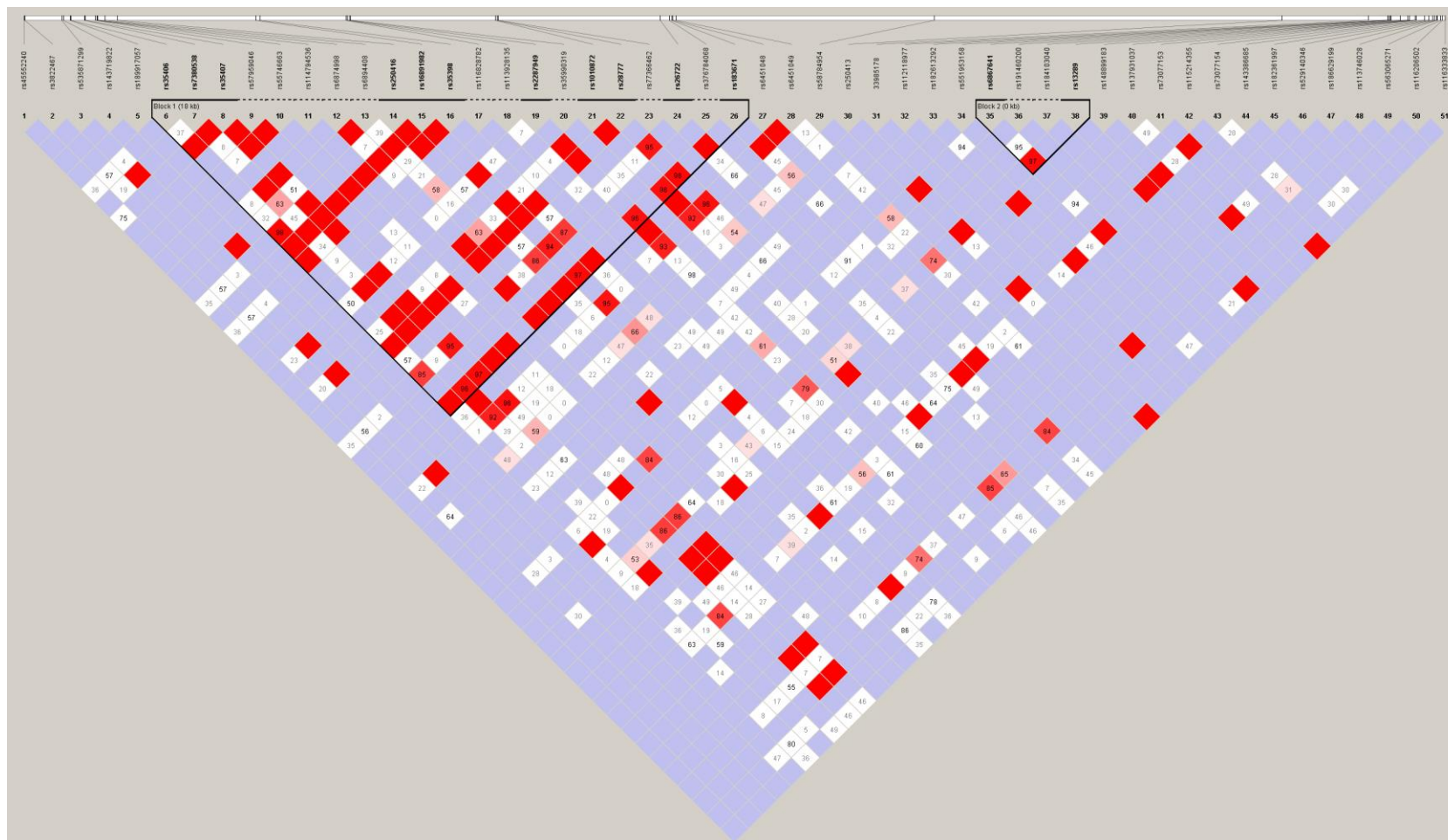


FIGURA 8. Representação do desequilíbrio de ligação para os pontos de variação do gene **SLC45A2**, considerando a região promotora no lado direito da imagem. O bloco em desequilíbrio está delimitado em preto. Vermelho indica LD completo forte [$LOD (Log\ of\ the\ Odds) \geq 2$ e $D' = 1$]; azul indica LD completo fraco [$LOD < 2$ e $D' = 1$]; tons de rosa indicam LD moderado [$LOD \geq 2$ e $D' < 1$]; branco indica ausência de LD [$LOD < 2$ e $D' < 1$]. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). Valores de D' diferentes de 1 estão representados na forma de porcentagem no interior dos quadros com tons de rosa ou brancos.



FIGURA 9. Representação do desequilíbrio de ligação para os pontos de variação do gene *TYR*. O bloco em desequilíbrio está delimitado em preto. Vermelho indica LD completo forte [LOD (Log of the Odds) ≥ 2 e $D' = 1$]; azul indica LD completo fraco [$LOD < 2$ e $D' = 1$]; tons de rosa indicam LD moderado [$LOD \geq 2$ e $D' < 1$]; branco indica ausência de LD [$LOD < 2$ e $D' < 1$]. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). Valores de D' diferentes de 1 estão representados na forma de porcentagem no interior dos quadros com tons de rosa ou brancos.

6.4 Associações fenotípicas

Com a finalidade de apresentar apenas associações fenotípicas com maior significância estatística, as Tabelas 13 e 14 mostram apenas associações envolvendo 28 variantes (das 58 avaliadas) do gene *SLC45A2* e 15 (das 42 avaliadas) do gene *TYR*, que respeitem simultaneamente os seguintes critérios: valores de probabilidade obtidos pelo teste exato de Fisher bilateral menores ou iguais a 0,05, valores de *Odds Ratio* (OR) maiores ou iguais a 10 e limite inferior do intervalo de confiança (IC) maiores que um.

Tabela 13: Alelos e genótipos de SNPs/Indels do gene *SLC45A2* associados com as características fenotípicas avaliadas.

SNPs/Indels	Alelo/ Genótipo	Característica	Valor de p	<i>Odds Ratio</i> (IC 95%)
rs45552240	T/G	cabelos cast claro x todos	0,0193	14,3793 (1,4696 - 140,6928)
	G	cabelos cast claro x todos	0,0197	14,0420 (1,4481 - 136,1646)
rs146922924	C/C	pele clara x pele escura	0,0001	27,1915 (3,3560 - 220,3131)
	C	pele clara x pele escura	0,0001	24,9029 (3,1166 - 198,9831)
	C/C	pele clara x pele média	0,0048	11,0938 (1,4006 - 87,8686)
	C	pele clara x pele média	0,0052	10,7143 (1,3623 - 84,2666)
	C/C	pele clara x todos	0,0003	15,4171 (2,0390 - 116,5719)
	C	pele clara x todos	0,0003	14,6748 (1,9529 - 110,2713)
	C/CA	pele escura x pele clara	0,0001	27,1915 (3,3560 - 220,3131)
	CA	pele escura x pele clara	0,0001	24,9029 (3,1166 - 198,9831)
	C/CA	pele média x pele clara	0,0048	11,0938 (1,4006 - 87,8686)
	CA	pele média x pele clara	0,0052	10,7143 (1,3623 - 84,2666)
rs35406	A/G	cabelos cast escuro x loiros	0,0005	14,5000 (1,9171 - 109,6731)
	G	cabelos cast escuro x loiros	0,0001	15,7907 (2,1324 - 116,9318)
	A/G	cabelos pretos x loiros	0,0002	16,5000 (2,1484 - 126,7248)
	G	cabelos pretos x loiros	0,0001	15,7297 (2,0935 - 118,1851)
	A/G	cabelos cast claro x loiros	0,0111	10,5600 (1,3126 - 84,9573)
	G	cabelos cast claro x loiros	0,0039	11,5688 (1,4859 - 90,0693)
	A/A	cabelos loiros x cast escuro	0,0002	16,5161 (2,1896 - 124,5792)
	A	cabelos loiros x cast escuro	0,0001	15,7907 (2,1324 - 116,9318)
	A/A	cabelos loiros x pretos	0,0001	17,5238 (2,2854 - 134,3689)
	A	cabelos loiros x pretos	0,0001	15,7297 (2,0935 - 118,1851)
	A/A	cabelos loiros x cast claro	0,0057	11,7551 (1,4708 - 93,9488)
	A	cabelos loiros x cast claro	0,0039	11,5688 (1,4859 - 90,0693)
	A/A	cabelos loiros x todos	0,0002	14,6368 (1,9832 - 108,0224)
	A	cabelos loiros x todos	0,0001	13,9116 (1,9107 - 101,2888)
rs7380538	G/G	pele clara x pele escura	6,55 x 10⁻¹²	19,9333 (7,5448 - 52,6637)
	G	pele clara x pele escura	6,63 x 10⁻¹²	15,6667 (6,2762 - 39,1069)
	A/G	pele escura x pele clara	1,25 x 10⁻¹⁰	17,2500 (6,5147 - 45,6754)
	A	pele escura x pele clara	6,63 x 10⁻¹²	15,6667 (6,2762 - 39,1069)
	A/A	pele escura x todos	0,0267	26,1009 (1,2359 - 551,2192)
	A/G	cabelos pretos x loiros	6,85 x 10⁻⁰⁶	24,4211 (3,2070 - 185,9645)
	A	cabelos pretos x loiros	1,85 x 10⁻⁰⁶	23,0288 (3,0971 - 171,2306)
	G/G	cabelos loiros x pretos	1,50 x 10⁻⁰⁶	27,0545 (3,5582 - 205,7092)
	G	cabelos loiros x pretos	1,85 x 10⁻⁰⁶	23,0288 (3,0971 - 171,2306)
	G/G	cabelos loiros x todos	0,0014	11,4383 (1,5455 - 84,6542)
	G	cabelos loiros x todos	0,0015	10,7366 (1,4697 - 78,4362)

rs35407	G/G	pele clara x pele escura	7,10 x 10⁻¹⁸	23,7931 (10,1482 - 55,7847)
	G	pele clara x pele escura	6,25 x 10⁻²⁵	13,4471 (7,9405 - 22,7722)
	A/A	pele escura x pele clara	3,05 x 10⁻¹¹	20,8500 (7,3893 - 58,8313)
	A	pele escura x pele clara	6,25 x 10⁻²⁵	13,4471 (7,9405 - 22,7722)
	A/A	olhos cast escuro x verdes	2,25 x 10⁻⁰⁵	11,0092 (2,5704 - 47,1525)
	A/A	olhos cast escuro x azuis	0,0007	13,2110 (1,7529 - 99,5653)
	A/G	cabelos cast escuro x loiros	2,83 x 10⁻⁰⁶	10,5676 (3,1162 - 35,8359)
	A	cabelos cast escuro x loiros	4,00 x 10⁻¹⁰	14,6296 (4,4961 - 47,6025)
	A/A	cabelos pretos x loiros	1,71 x 10⁻⁰⁷	53,4425 (3,1843 - 896,9244)
	A/G	cabelos pretos x loiros	1,53 x 10⁻⁰⁶	12,1389 (3,5022 - 42,0745)
	A	cabelos pretos x loiros	7,16 x 10⁻²²	41,9369 (12,7799 - 137,6146)
	A	cabelos pretos x ruivos	1,66 x 10⁻⁰⁵	13,2432 (3,0010 - 58,4426)
	G/G	cabelos loiros x cast escuro	3,33 x 10⁻⁰⁹	16,6111 (4,9067 - 56,2354)
	G	cabelos loiros x cast escuro	4,00 x 10⁻¹⁰	14,6296 (4,4961 - 47,6025)
	G/G	cabelos loiros x pretos	2,08 x 10⁻¹⁷	57,9259 (16,1339 - 207,9728)
	G	cabelos loiros x pretos	7,16 x 10⁻²²	41,9369 (12,7799 - 137,6146)
	G/G	cabelos loiros x todos	1,79 x 10⁻¹¹	18,4697 (5,6161 - 60,7409)
G	cabelos loiros x todos	1,36 x 10⁻¹³	17,8763 (5,5945 - 57,1212)	
G/G	cabelos ruivos x pretos	1,20 x 10⁻⁰⁴	17,0000 (3,3715 - 85,7187)	
G	cabelos ruivos x pretos	1,66 x 10⁻⁰⁵	13,2432 (3,0010 - 58,4426)	
rs57959046	C/C	pele clara x pele escura	3,60 x 10⁻⁰⁶	34,9556 (4,3918 - 278,2231)
	C	pele clara x pele escura	1,48 x 10⁻⁰⁶	34,4400 (4,4219 - 268,2369)
	C/C	pele clara x pele média	0,0013	13,4063 (1,7192 - 104,5441)
	C	pele clara x pele média	0,0014	12,8507 (1,6597 - 99,5022)
	C/C	pele clara x todos	2,31 x 10⁻⁰⁵	19,0116 (2,5363 - 142,5058)
	C	pele clara x todos	1,74 x 10⁻⁰⁵	18,7174 (2,5171 - 139,1815)
	C/A	pele escura x pele clara	1,36 x 10⁻⁰⁵	31,087 (3,8748 - 249,4078)
	A	pele escura x pele clara	1,48 x 10⁻⁰⁶	34,4400 (4,4219 - 268,2369)
	C/A	pele média x pele clara	0,0013	13,4063 (1,7192 - 104,5441)
A	pele média x pele clara	0,0014	12,8507 (1,6597 - 99,5022)	
rs55746663	A/A	pele clara x pele escura	9,32 x 10⁻⁰⁷	39,0000 (4,9321 - 308,3854)
	A	pele clara x pele escura	4,10 x 10⁻⁰⁷	37,6869 (4,8675 - 291,795)
	A/A	pele clara x pele média	0,0013	13,4063 (1,7192 - 104,5441)
	A	pele clara x pele média	0,0014	12,8507 (1,6597 - 99,5022)
	A/A	pele clara x todos	2,33 x 10⁻⁰⁵	19,9535 (2,6665 - 149,3128)
	A	pele clara x todos	1,71 x 10⁻⁰⁵	19,5504 (2,6334 - 145,1449)
	A/G	pele escura x pele clara	3,60 x 10⁻⁰⁶	34,9556 (4,3918 - 278,2231)
	G	pele escura x pele clara	4,10 x 10⁻⁰⁷	37,6869 (4,8675 - 291,795)
	A/G	pele média x pele clara	0,0013	13,4063 (1,7192 - 104,5441)
G	pele média x pele clara	0,0014	12,8507 (1,6597 - 99,5022)	
rs6874998	T/T	pele clara x pele escura	0,0057	24,7714 (1,3112 - 467,9935)
	T	pele clara x pele escura	0,0059	23,9309 (1,2778 - 448,1989)
	T/C	pele escura x pele clara	0,0057	24,7714 (1,3112 - 467,9935)
	C	pele escura x pele clara	0,0059	23,9309 (1,2778 - 448,1989)
rs6894408	C/C	pele clara x pele escura	2,58 x 10⁻⁰⁵	50,6495 (2,8699 - 893,8886)
	C	pele clara x pele escura	3,14 x 10⁻⁰⁵	46,9330 (2,6853 - 820,2859)
	C/C	pele clara x pele média	0,0031	18,5396 (1,0597 - 324,3435)
	C	pele clara x pele média	0,0033	17,9982 (1,0339 - 313,3240)
	C/C	pele clara x todos	0,0001	26,4183 (1,5716 - 444,0963)
	C	pele clara x todos	0,0002	25,2869 (1,5108 - 423,2455)
	C/G	pele escura x pele clara	2,58 x 10⁻⁰⁵	50,6495 (2,8699 - 893,8886)
	G	pele escura x pele clara	3,14 x 10⁻⁰⁵	46,9330 (2,6853 - 820,2859)
	C/G	pele média x pele clara	0,0031	18,5396 (1,0597 - 324,3435)
G	pele média x pele clara	0,0033	17,9982 (1,0339 - 313,324)	
rs250416	C/A	cabelos cast escuro x loiros	0,0002	16,1720 (2,1431 - 122,0352)
	A	cabelos cast escuro x loiros	3,68 x 10⁻⁰⁵	17,0283 (2,3039 - 125,8579)
	C/A	cabelos pretos x loiros	0,0002	16,1563 (2,1027 - 124,1365)
	A	cabelos pretos x loiros	0,0003	15,4054 (2,0499 - 115,7723)

	C/A	cabelos cast claro x loiros	0,0057	11,7500 (1,4690 - 93,9834)
	A	cabelos cast claro x loiros	0,0073	10,5556 (1,3473 - 82,6977)
	C/C	cabelos loiros x cast escuro	4,35 x 10⁻⁰⁵	18,2778 (2,4275 - 137,6212)
	C	cabelos loiros x cast escuro	3,68 x 10⁻⁰⁵	17,0283 (2,3039 - 125,8579)
	C/C	cabelos loiros x pretos	0,0001	17,1587 (2,2368 - 131,6245)
	C	cabelos loiros x pretos	0,0003	15,4054 (2,0499 - 115,7723)
	C/C	cabelos loiros x cast claro	0,0057	11,7500 (1,4690 - 93,9834)
	C	cabelos loiros x cast claro	0,0073	10,5556 (1,3473 - 82,6977)
	C/C	cabelos loiros x todos	0,0001	15,2374 (2,0648 - 112,4451)
	C	cabelos loiros x todos	0,0001	14,1089 (1,9380 - 102,7123)
rs16891982	G/G	pele clara x pele escura	6,27 x 10⁻²⁴	171,2857 (22,8580 - 1283,5172)
	G	pele clara x pele escura	9,54 x 10⁻³⁵	22,6646 (12,9041 - 39,8078)
	G/G	pele clara x pele média	7,08 x 10⁻¹⁹	10,0978 (5,8536 - 17,4193)
	G/G	pele clara x todos	5,42 x 10⁻²⁸	14,8387 (8,7276 - 25,2288)
	C/C	pele escura x pele clara	6,08 x 10⁻¹⁷	30,2468 (11,9371 - 76,6407)
	C	pele escura x pele clara	9,54 x 10⁻³⁵	22,6646 (12,9041 - 39,8078)
	G/G	pele média x pele escura	0,0001	16,9626 (2,2596 - 127,3389)
	C/C	olhos cast escuro x azuis	6,12 x 10⁻⁰⁶	21,6774 (2,8916 - 162,5095)
	C/G	cabelos cast escuro x loiros	1,20 x 10⁻⁰⁷	13,2736 (3,9201 - 44,9449)
	C/C	cabelos pretos x cast claro	8,85 x 10⁻⁰⁸	10,6909 (3,9024 - 29,2884)
	C/C	cabelos pretos x loiros	1,21 x 10⁻⁰⁸	22,4318 (5,1222 - 98,2367)
	C/G	cabelos pretos x loiros	1,53 x 10⁻⁰⁶	12,1389 (3,5022 - 42,0745)
	C	cabelos pretos x loiros	1,59 x 10⁻²⁶	31,7200 (13,7453 - 73,2000)
	C/C	cabelos pretos x ruivos	0,0021	21,9663 (1,2549 - 384,5092)
	C	cabelos pretos x ruivos	2,82 x 10⁻⁰⁷	15,4533 (4,3775 - 54,5527)
	G/G	cabelos cast claro x pretos	1,29 x 10⁻¹²	20,5556 (7,7474 - 54,5387)
	G/G	cabelos loiros x cast escuro	2,21 x 10⁻¹¹	16,2000 (5,9886 - 43,8232)
	G/G	cabelos loiros x pretos	2,88 x 10⁻²³	117,3333 (33,8689 - 406,4822)
	G	cabelos loiros x pretos	1,59 x 10⁻²⁶	31,7200 (13,7453 - 73,2000)
	G/G	cabelos loiros x todos	8,09 x 10⁻¹⁴	17,0667 (6,5593 - 44,4060)
	G	cabelos loiros x todos	7,07 x 10⁻¹⁵	10,7925 (4,9183 - 23,6825)
	G/G	cabelos ruivos x pretos	2,84 x 10⁻⁰⁶	35,5556 (7,4349 - 170,0346)
G	cabelos ruivos x pretos	2,82 x 10⁻⁰⁷	15,4533 (4,3775 - 54,5527)	
rs35398	G/T	cabelos cast escuro x loiros	0,0005	14,5000 (1,9171 - 109,6731)
	T	cabelos cast escuro x loiros	0,0001	15,7907 (2,1324 - 116,9318)
	G/T	cabelos pretos x loiros	0,0002	16,5000 (2,1484 - 126,7248)
	T	cabelos pretos x loiros	0,0001	15,7297 (2,0935 - 118,1851)
	G/T	cabelos cast claro x loiros	0,0057	11,7551 (1,4708 - 93,9488)
	T	cabelos cast claro x loiros	0,0073	10,5818 (1,3512 - 82,8737)
	G/G	cabelos loiros x cast escuro	0,0002	16,5161 (2,1896 - 124,5792)
	G	cabelos loiros x cast escuro	0,0001	15,7907 (2,1324 - 116,9318)
	G/G	cabelos loiros x pretos	0,0001	17,5238 (2,2854 - 134,3689)
	G	cabelos loiros x pretos	0,0001	15,7297 (2,0935 - 118,1851)
	G/G	cabelos loiros x cast claro	0,0057	11,7551 (1,4708 - 93,9488)
	G	cabelos loiros x cast claro	0,0073	10,5818 (1,3512 - 82,8737)
	G/G	cabelos loiros x todos	0,0002	14,6368 (1,9832 - 108,0224)
G	cabelos loiros x todos	0,0001	13,6941 (1,8805 - 99,7234)	
rs113928135	G/A	cabelos pretos x todos	0,0156	12,3415 (1,3601 - 111,9866)
	A	cabelos pretos x todos	0,0160	12,0714 (1,3399 - 108,7534)
rs2287949	C/C	pele clara x pele escura	8,89 x 10⁻⁰⁹	15,4444 (5,4247 - 43,9711)
	C	pele clara x pele escura	3,33 x 10⁻⁰⁸	12,3043 (4,4913 - 33,7091)
	T/C	pele escura x pele clara	8,89 x 10⁻⁰⁹	15,4444 (5,4247 - 43,9711)
	T	pele escura x pele clara	3,33 x 10⁻⁰⁸	12,3043 (4,4913 - 33,7091)
	T/C	cabelos pretos x loiros	2,91 x 10⁻⁰⁵	20,8000 (2,7235 - 158,8556)
	T	cabelos pretos x loiros	0,0001	17,2740 (2,3059 - 129,4033)
	C/C	cabelos loiros x pretos	2,91 x 10⁻⁰⁵	20,8000 (2,7235 - 158,8556)
C	cabelos loiros x pretos	0,0001	17,2740 (2,3059 - 129,4033)	
rs1010872	C/C	pele clara x pele escura	6,83 x 10⁻¹²	22,4194 (7,9552 - 63,1824)

	C	pele clara x pele escura	2,01 x 10⁻¹¹	17,1116 (6,3770 - 45,9161)
	A/C	pele escura x pele clara	3,05 x 10⁻¹¹	20,8500 (7,3893 - 58,8313)
	A	pele escura x pele clara	2,01 x 10⁻¹¹	17,1116 (6,3770 - 45,9161)
	A/C	cabelos pretos x loiros	6,85 x 10⁻⁰⁶	24,4211 (3,2070 - 185,9645)
	A	cabelos pretos x loiros	6,33 x 10⁻⁰⁶	21,3262 (2,8631 - 158,8528)
	C/C	cabelos loiros x pretos	3,15 x 10⁻⁰⁶	25,7143 (3,3795 - 195,6575)
	C	cabelos loiros x pretos	6,33 x 10⁻⁰⁶	21,3262 (2,8631 - 158,8528)
	C/C	cabelos loiros x todos	0,0015	10,9367 (1,4769 - 80,9906)
	C	cabelos loiros x todos	0,0023	10,1233 (1,3845 - 74,0222)
rs28777	A/A	pele clara x pele escura	4,91 x 10⁻¹⁹	30,7611 (12,0348 - 78,6257)
	A	pele clara x pele escura	1,01 x 10⁻²⁶	14,8810 (8,7413 - 25,3329)
	C/C	pele escura x pele clara	4,77 x 10⁻¹²	20,4713 (7,73 - 54,2135)
	C	pele escura x pele clara	1,01 x 10⁻²⁶	14,8810 (8,7413 - 25,3329)
	C/C	olhos cast escuro x azuis	0,0004	14,8077 (1,9667 - 111,4887)
	C/A	cabelos cast escuro x loiros	3,05 x 10⁻⁰⁶	10,3378 (3,0462 - 35,0835)
	C	cabelos cast escuro x loiros	5,21 x 10⁻¹¹	15,9697 (4,9116 - 51,9244)
	C/C	cabelos pretos x loiros	2,98 x 10⁻⁰⁸	61,8952 (3,6909 - 1037,9587)
	C/A	cabelos pretos x loiros	7,20 x 10⁻⁰⁷	12,7174 (3,665 - 44,1283)
	C	cabelos pretos x loiros	2,83 x 10⁻²⁴	50,0769 (15,2257 - 164,7022)
	C	cabelos pretos x ruivos	2,83 x 10⁻⁰⁶	16,1538 (3,6549 - 71,3970)
	A/A	cabelos cast claro x pretos	4,45 x 10⁻¹⁰	10,5495 (4,7766 - 23,2992)
	A/A	cabelos loiros x cast escuro	1,03 x 10⁻⁰⁹	17,8947 (5,2802 - 60,6461)
	A	cabelos loiros x cast escuro	5,21 x 10⁻¹¹	15,9697 (4,9116 - 51,9244)
	A/A	cabelos loiros x pretos	5,07 x 10⁻²⁰	83,0769 (22,4278 - 307,7333)
	A	cabelos loiros x pretos	2,83 x 10⁻²⁴	50,0769 (15,2257 - 164,7022)
	A/A	cabelos loiros x todos	6,65 x 10⁻¹²	20,0806 (6,0990 - 66,1141)
	A	cabelos loiros x todos	1,18 x 10⁻¹⁴	19,3641 (6,0595 - 61,8816)
A/A	cabelos ruivos x pretos	1,58 x 10⁻⁰⁵	24,9231 (4,8246 - 128,7495)	
A	cabelos ruivos x pretos	2,83 x 10⁻⁰⁶	16,1538 (3,6549 - 71,3970)	
rs141348470	A/A	pele clara x pele escura	8,57 x 10⁻⁰⁸	10,1250 (4,1365 - 24,7834)
	AAAC/A	pele escura x pele clara	8,57 x 10⁻⁰⁸	10,1250 (4,1365 - 24,7834)
	AAAC/A	cabelos pretos x loiros	0,0001	10,5254 (2,3789 - 46,5696)
	A/A	cabelos loiros x pretos	0,0001	10,5254 (2,3789 - 46,5696)
rs26722	C/T	cabelos cast escuro x loiros	0,0002	15,8298 (2,0969 - 119,5036)
	T	cabelos cast escuro x loiros	0,0002	14,7512 (1,9888 - 109,4109)
	C/T	cabelos pretos x loiros	0,0002	16,5 (2,1484 - 126,7248)
	T	cabelos pretos x loiros	2,21 x 10⁻⁰⁵	18,8611 (2,5241 - 140,9358)
	C/T	cabelos cast claro x loiros	0,0014	14,2979 (1,8074 - 113,1087)
	T	cabelos cast claro x loiros	0,0021	12,5741 (1,6233 - 97,4009)
	C/C	cabelos loiros x cast escuro	0,0002	16,5161 (2,1896 - 124,5792)
	C	cabelos loiros x cast escuro	0,0002	14,7512 (1,9888 - 109,4109)
	C/C	cabelos loiros x pretos	0,0001	19,6721 (2,5727 - 150,4214)
	C	cabelos loiros x pretos	2,21 x 10⁻⁰⁵	18,8611 (2,5241 - 140,9358)
	C/C	cabelos loiros x cast claro	0,0014	14,2979 (1,8074 - 113,1087)
	C	cabelos loiros x cast claro	0,0021	12,5741 (1,6233 - 97,4009)
	C/C	cabelos loiros x todos	0,0001	16,3687 (2,2202 - 120,6801)
C	cabelos loiros x todos	0,0001	15,0119 (2,0635 - 109,2093)	
rs376784068	T/A	pele escura x todos	0,0270	25,9174 (1,2272 - 547,3528)
	A	pele escura x todos	0,0272	25,5430 (1,2179 - 535,6942)
rs183671	G/G	pele clara x pele escura	3,6 x 10⁻¹⁹	33,6909 (12,4236 - 91,3646)
	G	pele clara x pele escura	6,35 x 10⁻²⁵	12,8778 (7,6662 - 21,6323)
	T/T	pele escura x pele clara	1,41 x 10⁻¹⁰	15,5530 (6,1700 - 39,2054)
	T	pele escura x pele clara	6,35 x 10⁻²⁵	12,8778 (7,6662 - 21,6323)
	T/T	olhos cast escuro x azuis	0,0004	14,1981 (1,8851 - 106,9344)
	T/G	cabelos cast escuro x loiros	3,05 x 10⁻⁰⁶	10,3378 (3,0462 - 35,0835)
	T/T	cabelos pretos x loiros	6,66 x 10⁻⁰⁷	29,2642 (3,8518 - 222,3360)
	T/G	cabelos pretos x loiros	2,96 x 10⁻⁰⁷	13,6667 (3,9433 - 47,3660)
T	cabelos pretos x loiros	2,39 x 10⁻²²	29,9600 (11,5685 - 77,5904)	

	T	cabelos pretos x ruivos	1,55 x 10⁻⁰⁵	10,4256 (2,9690 - 36,6100)
	G/G	cabelos cast claro x pretos	6,52 x 10⁻¹⁰	10,6515 (4,763 - 23,8199)
	G/G	cabelos loiros x cast escuro	1,29 x 10⁻⁰⁸	12,7069 (4,3056 - 37,5016)
	G/G	cabelos loiros x pretos	1,22 x 10⁻¹⁹	67,8333 (20,6051 - 223,3113)
	G	cabelos loiros x pretos	2,39 x 10⁻²²	29,9600 (11,5685 - 77,5904)
	G/G	cabelos loiros x todos	2,69 x 10⁻¹¹	15,3636 (5,3772 - 43,8966)
	G	cabelos loiros x todos	9,58 x 10⁻¹³	11,5352 (4,6171 - 28,8193)
	G/G	cabelos ruivos x pretos	0,0001	16,4444 (3,8176 - 70,8358)
	G	cabelos ruivos x pretos	1,55 x 10⁻⁰⁵	10,4256 (2,9690 - 36,6100)
rs6451048	G/G	pele clara x pele escura	0,0059	24,4286 (1,2930 - 461,5453)
	G	pele clara x pele escura	0,0062	23,5991 (1,2600 - 441,9988)
	A/G	pele escura x pele clara	0,0059	24,4286 (1,2930 - 461,5453)
	A	pele escura x pele clara	0,0062	23,5991 (1,2600 - 441,9988)
	A/G	pele escura x pele média	0,0062	24,0857 (1,2747 - 455,0972)
	A	pele escura x pele média	0,0064	23,2673 (1,2422 - 435,7987)
	A/G	pele escura x todos	0,0007	48,4286 (2,5691 - 912,9152)
	A	pele escura x todos	0,0007	46,8249 (2,5029 - 876,0044)
	G/G	pele média x pele escura	0,0062	24,0857 (1,2747 - 455,0972)
	G	pele média x pele escura	0,0064	23,2673 (1,2422 - 435,7987)
	A/G	cabelos pretos x todos	0,0040	27,5455 (1,4675 - 517,0462)
	A	cabelos pretos x todos	0,0041	26,9466 (1,4433 - 503,1047)
rs6451049	C/C	pele clara x pele escura	1,50 x 10⁻⁰⁷	11,5541 (4,3050 - 31,0099)
	T/C	pele escura x pele clara	1,50 x 10⁻⁰⁷	11,5541 (4,3050 - 31,0099)
rs58784954	G/A	pele escura x todos	0,0151	16,0189 (1,6350 - 156,9412)
	A	pele escura x todos	0,0154	15,6055 (1,6083 - 151,4203)
rs250413	G/A	cabelos pretos x loiros	0,0041	17,8163 (1,0348 - 306,7334)
	G/G	cabelos loiros x pretos	0,0041	17,8163 (1,0348 - 306,7334)
rs551953158	C/T	pele escura x todos	0,0267	26,1009 (1,2359 - 551,2192)
	T	pele escura x todos	0,0269	25,7240 (1,2266 - 539,4858)
rs73077153	C/T	pele escura x todos	0,0267	26,1009 (1,2359 - 551,2192)
	T	pele escura x todos	0,0269	25,7240 (1,2266 - 539,4858)
rs73077154	G/G	pele clara x pele escura	1,39 x 10⁻⁰⁶	21,4651 (4,6606 - 98,8614)
	G	pele clara x pele escura	2,34 x 10⁻⁰⁶	18,7778 (4,1642 - 84,6745)
	G/C	pele escura x pele clara	1,39 x 10⁻⁰⁶	21,4651 (4,6606 - 98,8614)
	C	pele escura x pele clara	2,34 x 10⁻⁰⁶	18,7778 (4,1642 - 84,6745)
	G/C	cabelos pretos x loiros	0,0042	18,1837 (1,0565 - 312,9698)
	G/G	cabelos loiros x pretos	0,0042	18,1837 (1,0565 - 312,9698)
rs143386685	G/A	cabelos pretos x todos	0,0158	21,3353 (1,0908 - 417,3158)
	A	cabelos pretos x todos	0,0160	21,0000 (1,0792 - 408,6275)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Considerando que cada um dos 58 SNPs possui apenas dois alelos, a informação de cada um dos dois torna-se redundante. Portanto, para alelos, foi considerado $\alpha_{\text{corrigido}} = 0,0009$ (isto é, $0,05/58$). Considerando que cada um dos 58 SNPs possui apenas três genótipos, foi considerado $\alpha_{\text{corrigido}} = 0,0003$ (isto é, $0,05/174$).

Tabela 14: Alelos e genótipos de SNPs/Indels do gene *TYR* associados com as características fenotípicas avaliadas.

SNPs/Indels	Alelo/ Genótipo	Característica	Valor de <i>p</i>	Odds Ratio (IC 95%)
rs75741664	C/C	pele clara x pele escura	0,0224	11,0000 (1,2017 - 100,6922)
	C	pele clara x pele escura	0,0231	10,6296 (1,1749 - 96,1677)
	C/T	pele escura x pele clara	0,0224	11,0000 (1,2017 - 100,6922)
	T	pele escura x pele clara	0,0231	10,6296 (1,1749 - 96,1677)
88907815	A/G	cabelos loiros x todos	0,0204	30,6842 (1,4505 - 649,0823)
	G	cabelos loiros x todos	0,0206	30,1813 (1,4380 - 633,4739)

rs142881018	CCTGA/CCTGA	pele clara x pele escura	0,0001	43,7879 (2,4559 - 780,7133)
	CCTGA	pele clara x pele escura	0,0001	41,0190 (2,3225 - 724,4532)
	CCTGA/CCTGA	pele clara x todos	0,0031	17,9164 (1,0470 - 306,5808)
	CCTGA	pele clara x todos	0,0034	17,3932 (1,0207 - 296,3758)
	CCTGA/C	pele escura x pele clara	0,0001	43,7879 (2,4559 - 780,7133)
	C	pele escura x pele clara	0,0001	41,019 (2,3225 - 724,4532)
rs61759512	CCTGA/C	pele escura x todos	0,0005	10,0000 (2,8212 - 35,4454)
	G/G	pele clara x pele escura	0,0232	10,8462 (1,1848 - 99,2924)
	G	pele clara x pele escura	0,0240	10,4815 (1,1585 - 94,8314)
	G/A	pele escura x pele clara	0,0232	10,8462 (1,1848 - 99,2924)
rs11824466	A	pele escura x pele clara	0,0240	10,4815 (1,1585 - 94,8314)
	C/C	pele clara x pele escura	0,0002	23,8333 (2,9057 - 195,4869)
	C	pele clara x pele escura	0,0001	25,0777 (3,1386 - 200,375)
	C	pele clara x todos	0,0033	10,6296 (1,3897 - 81,3041)
	C/T	pele escura x pele clara	0,0006	20,4286 (2,4514 - 170,2374)
	T	pele escura x pele clara	0,0001	25,0777 (3,1386 - 200,375)
rs55870815	C/T	pele escura x todos	0,0005	10,0000 (2,8212 - 35,4454)
	T/C	cabelos pretos x todos	0,0159	21,2515 (1,0865 - 415,6802)
rs1042602	C	cabelos pretos x todos	0,0161	20,9174 (1,0750 - 407,0224)
	A/A	pele clara x pele escura	0,0033	11,0000 (1,4509 - 83,3938)
rs11018530	G/G	pele escura x pele clara	2,69 x 10⁻⁰⁵	50,299 (2,8499 - 887,7312)
	G/G	pele escura x todos	3,89 x 10⁻⁰⁵	15,5556 (3,9854 - 60,7148)
	C/G	olhos cast escuro x azuis	0,0001	10,5326 (2,4366 - 45,5280)
	G	olhos cast escuro x azuis	5,48 x 10⁻⁰⁶	11,3556 (2,7171 - 47,4581)
	C/C	olhos azuis x cast escuro	4,96 x 10⁻⁰⁶	13,1548 (3,0478 - 56,7786)
	C	olhos azuis x cast escuro	5,48 x 10⁻⁰⁶	11,3556 (2,7171 - 47,4581)
rs67512950	A/A	pele escura x pele clara	0,0054	25,2524 (1,3363 - 477,1960)
	ACTT/ACTT	olhos azuis x cast escuro	2,58 x 10⁻⁰⁵	11,2778 (2,6121 - 48,6914)
rs12804012	A/A	pele escura x pele clara	1,67 x 10⁻⁰⁶	64,8065 (3,7250 - 1127,3817)
	A/A	pele escura x todos	1,40 x 10⁻⁰⁶	20,2899 (5,3804 - 76,5139)
	G/G	olhos azuis x cast escuro	3,72 x 10⁻⁰⁶	10,2857 (3,0221 - 35,0073)
	G/G	cabelos ruivos x pretos	0,0108	10,0000 (1,2262 - 81,5505)
rs75375648	T/T	pele clara x pele escura	0,0015	30,6505 (1,6655 - 564,0599)
	T	pele clara x pele escura	0,0017	29,3163 (1,6074 - 534,6844)
	T/C	pele escura x pele clara	0,0015	30,6505 (1,6655 - 564,0599)
	C	pele escura x pele clara	0,0017	29,3163 (1,6074 - 534,6844)
rs113438665	T/T	pele clara x pele escura	0,0015	30,6505 (1,6655 - 564,0599)
	T	pele clara x pele escura	0,0017	29,3163 (1,6074 - 534,6844)
	T/C	pele escura x pele clara	0,0015	30,6505 (1,6655 - 564,0599)
	C	pele escura x pele clara	0,0017	29,3163 (1,6074 - 534,6844)
rs1393350	A/A	olhos cast claro x cast escuro	0,0155	10,8824 (1,2473 - 94,9443)
	A/A	olhos azuis x cast escuro	0,0253	13,0588 (1,3176 - 129,4282)
	A/A	cabelos cast claro x cast escuro	0,0019	16,0741 (1,9303 - 133,8508)
	A/A	cabelos cast claro x pretos	0,0091	11,0185 (1,3188 - 92,0605)
rs1126809	A/A	olhos cast claro x cast escuro	0,0158	10,8209 (1,2399 - 94,4382)
	A/A	olhos azuis x cast escuro	0,0249	13,1818 (1,3288 - 130,7600)
	A/A	cabelos cast claro x cast escuro	0,0056	13,4444 (1,5800 - 114,4034)
rs28521275	C/C	cabelos ruivos x pretos	0,0014	10,8000 (2,6804 - 43,516)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Considerando que cada um dos 42 SNPs possui apenas dois alelos, a informação de cada um dos dois torna-se redundante. Portanto, para alelos, foi considerado $\alpha_{\text{corrigido}} = 0,0012$ (isto é, $0,05/42$). Considerando que cada um dos 42 SNPs possui apenas três genótipos, foi considerado $\alpha_{\text{corrigido}} = 0,0004$ (isto é, $0,05/126$).

Com relação às associações encontradas envolvendo haplótipos, divididos de acordo com as regiões gênicas, foram mantidos os mesmos critérios citados acima para selecionar as associações mais fortes.

Tabela 15: Haplótipos formados por SNPs/indels das regiões avaliadas para o gene *SLC45A2* associados com as características fenóticas avaliadas.

Haplótipos	Característica	Valor de <i>p</i>	Odds Ratio (IC 95%)
Promotora			
promo3	cabelos castanhos x loiros	0,0003	12,8864 (1,7513 - 94,8226)
	cabelos pretos x loiros	0,0001	15,8333 (2,1065 - 119,0094)
promo15	cabelos pretos x outros	0,0174	20,2387 (1,0399 - 393,8810)
promo17	pele escura x outras	0,0245	27,2439 (1,2984 - 571,6279)
Codificante isoforma 3			
Iso3cds1	pele clara x pele escura	2,82 x 10⁻²⁵	17,2252 (9,6147 - 30,8598)
	cabelos loiros x castanhos	3,07 x 10⁻⁰⁹	12,6803 (3,9235 - 40,9818)
	cabelos loiros x outros	5,98 x 10⁻¹⁴	18,6224 (5,8151 - 59,6361)
	cabelos loiros x pretos	3,36 x 10⁻²³	50,9371 (15,3526 - 169,0004)
	cabelos ruivos x pretos	1,57 x 10⁻⁰⁵	10,6352 (3,0039 - 37,6531)
Iso3cds2	pele escura x pele clara	4,35 x 10⁻¹⁶	33,4237 (11,3209 - 98,6802)
	pele media x pele clara	7,73 x 10⁻⁰⁸	10,0628 (3,5119 - 28,8329)
	cabelos pretos x loiros	1,09 x 10⁻⁰⁷	29,2778 (3,9315 - 218,0323)
	olhos cast escuro x azuis	0,0001	15,7081 (2,1282 - 115,9398)
	olhos cast escuro x verdes	8,96 x 10⁻⁰⁷	13,3636 (3,1910 - 55,9657)
Iso3cds3	cabelos castanhos x loiros	0,0005	19,5310 (1,1847 - 321,9829)
Iso3cds4	pele escura x pele clara	3,89 x 10⁻⁰⁶	10,2541 (3,5776 - 29,3900)
	cabelos pretos x loiros	0,0002	16,1405 (2,1321 - 122,186)
Iso3cds5	cabelos pretos x outros	0,0128	23,2079 (1,1916 - 452,0176)
Iso3cds7	pele escura x pele clara	0,0138	22,6374 (1,1578 - 442,6151)
Iso3cds11	cabelos pretos x castanhos	0,0268	10,0000 (1,1078 - 90,2664)
	cabelos pretos x outros	0,0121	13,3623 (1,4813 - 120,5401)
Iso3cds13	pele escura x pele clara	0,0032	29,4497 (1,5696 - 552,5619)
	pele escura x pele media	0,0041	27,3195 (1,4558 - 512,6883)
	pele escura x outras	0,0004	56,7160 (3,0262 - 1062,9445)
	cabelos pretos x castanhos	0,0070	22,5162 (1,2043 - 420,9925)
	cabelos pretos x outros	0,0030	30,0542 (1,6082 - 561,6641)
Codificante isoforma 2			
Iso2cds1	cabelos loiros x pretos	6,34 x 10⁻²³	26,1905 (11,6682 - 58,7870)
	cabelos ruivos x pretos	4,28 x 10⁻⁰⁶	10,7143 (3,4205 - 33,5607)
	pele clara x pele escura	3,65 x 10⁻²⁹	22,6306 (12,1594 - 42,1195)
Iso2cds2	cabelos pretos x loiros	9,76 x 10⁻⁰⁸	29,9074 (4,0169 - 222,6729)
	olhos cast escuro x azuis	0,0001	15,3865 (2,0836 - 113,6208)
	olhos cast escuro x verdes	4,52 x 10⁻⁰⁷	13,7295 (3,2788 - 57,4897)
Iso2cds4	pele escura x pele clara	4,35 x 10⁻¹⁶	33,4237 (11,3209 - 98,6802)
	cabelos pretos x loiros	0,0001	16,4876 (2,1784 - 124,7869)
Iso2cds6	pele escura x pele clara	3,89 x 10⁻⁰⁶	10,2541 (3,5776 - 29,3900)
	cabelos pretos x outros	0,0132	22,9068 (1,1761 - 446,1618)
Iso2cds7	pele escura x pele clara	0,0138	22,6374 (1,1578 - 442,6151)
Iso2cds8	olhos cast claro x outros	0,0373	10,5556 (1,0892 - 102,2972)
Iso2cds11	cabelos pretos x outros	0,0125	13,1884 (1,4619 - 118,9742)
Iso2cds12	cabelos pretos x castanhos	0,0074	21,9964 (1,1764 - 411,2910)
	cabelos pretos x outros	0,0031	29,6643 (1,5873 - 554,3880)
	pele escura x pele clara	0,0032	29,4497 (1,5696 - 552,5619)
	pele escura x pele media	0,0044	26,6805 (1,4216 - 500,7262)

	pele escura x outras	0,0004	56,0769 (2,9921 - 1050,9824)
Codificante isoforma 1			
Iso1cds1	cabelos loiros x pretos	1,94 x 10⁻²²	22,9297 (10,765 - 48,8408)
	cabelos ruivos x preto	1,23 x 10⁻⁰⁶	12,2838 (3,9011 - 38,6788)
	pele clara x pele escura	1,08 x 10⁻²⁴	17,3875 (9,4198 - 32,0949)
Iso1cds3	cabelos pretos x loiros	1,93 x 10⁻⁰⁵	32,6962 (1,952 - 547,6619)
	olhos cast escuro x verdes	0,0004	13,6903 (1,8351 - 102,1338)
	pele escura x pele clara	7,08 x 10⁻⁰⁹	55,0000 (7,1113 - 425,3813)
	pele media x pele clara	3,99 x 10⁻⁰⁶	23,5043 (3,1307 - 176,4607)
Iso1cds5	cabelos pretos x loiros	0,0005	14,8487 (1,9519 - 112,9574)
	pele escura x pele clara	3,31 x 10⁻⁰⁶	15,1667 (4,1686 - 55,1813)
Iso1cds6	pele escura x pele clara	4,90 x 10⁻⁰⁵	28,9474 (3,565 - 235,0491)
	pele media x pele clara	0,0044	11,2705 (1,4324 - 88,6796)
Iso1cds7	cabelos pretos x castanhos	0,0005	33,8000 (1,8909 - 604,1793)
	cabelos pretos x outros	0,0001	45,1811 (2,5288 - 807,2252)
Iso1cds9	olhos cast claro x outros	0,0388	10,3577 (1,0688 - 100,3710)
Iso1cds10	pele escura x pele clara	0,0124	23,7485 (1,2142 - 464,5108)
Iso1cds13	cabelos pretos x castanhos	0,0065	23,0520 (1,2327 - 431,0760)
	cabelos pretos x outros	0,0027	30,8141 (1,6486 - 575,9515)
	pele escura x pele clara	0,0028	30,9130 (1,6469 - 580,2418)
	pele escura x pele media	0,0036	28,4534 (1,5156 - 534,1837)
	pele escura x outras	0,0003	59,3106 (3,1634 - 1112,0046)
3'UTR isoforma 2			
Iso2utr1	cabelos loiros x pretos	4,22 x 10⁻¹⁹	17,7273 (8,0766 - 38,9095)
	cabelos ruivos x pretos	3,19 x 10⁻⁰⁶	15,7576 (3,5659 - 69,6328)
	pele clara x pele escura	1,10 x 10⁻²²	11,2414 (6,7414 - 18,7451)
Iso2utr2	cabelos castanhos x loiros	0,0001	14,5500 (1,9824 - 106,7889)
	cabelos pretos x loiros	0,0001	15,9452 (2,122 - 119,8147)
Iso2utr3	olhos cast escuro x verdes	7,20 x 10⁻⁰⁶	19,2188 (2,6116 - 141,4315)
	cabelos pretos x loiros	7,98 x 10⁻⁰⁷	41,0707 (2,4801 - 680,1413)
	olhos cast escuro x azuis	0,0002	22,8947 (1,3908 - 376,8813)
	olhos mel x verdes	0,0405	10,5429 (1,0633 - 104,5386)
	pele escura x pele clara	1,89 x 10⁻⁰⁹	16,7528 (5,6020 - 50,0990)
Iso2utr4	cabelos pretos x loiros	2,09 x 10⁻⁰⁵	19,1268 (2,5595 - 142,9341)
	pele escura x pele clara	8,01 x 10⁻⁰⁹	11,7500 (4,6181 - 29,8962)
Iso2utr5	pele escura x pele clara	1,25 x 10⁻⁰⁶	35,1429 (4,5113 - 273,7634)
	pele media x pele clara	0,0013	13,0455 (1,6847 - 101,0152)
Iso2utr7	pele escura x pele clara	0,0056	24,3803 (1,3016 - 456,6675)
Iso2utr8	cabelos pretos x outros	0,0173	11,7349 (1,3024 - 105,736)
Iso2utr9	pele escura x outras	0,0264	26,0138 (1,2403 - 545,6262)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Para a região promotora foi considerado $\alpha_{\text{corrigido}} = 0,0024$ (isto é, 0,05/21 haplótipos). Para a região codificante da isoforma 3, foi considerado $\alpha_{\text{corrigido}} = 0,0021$ (isto é, 0,05/24 haplótipos). Para a região codificante da isoforma 2, foi considerado $\alpha_{\text{corrigido}} = 0,0018$ (isto é, 0,05/28 haplótipos). Para a região codificante da isoforma 1, foi considerado $\alpha_{\text{corrigido}} = 0,0014$ (isto é, 0,05/36 haplótipos). Para a região 3'UTR da isoforma 2, foi considerado $\alpha_{\text{corrigido}} = 0,005$ (isto é, 0,05/10 haplótipos).

Tabela 16: Haplótipos formados por SNPs/indels das regiões avaliadas para o gene *TYR* associados com as características fenotípicas avaliadas.

Haplótipos	Característica	Valor de p	Odds Ratio (IC 95%)
Promotora			
promo7	pele escura x pele clara	0,0011	33,0000 (1,8072 - 602,5785)
	pele escura x outras	0,0014	14,0659 (2,6881 - 73,6018)
promo8	pele escura x pele clara	0,0001	24,8182 (3,0614 - 201,1940)
	pele escura x pele media	0,0002	21,7273 (2,6789 - 176,2212)
	pele escura x outras	9,95 x 10⁻⁰⁶	23,2727 (4,8616 - 111,407)
promo9	pele escura x pele clara	0,0173	11,8696 (1,3099 - 107,5532)
promo10	pele escura x pele clara	0,0011	33,0000 (1,8072 - 602,5785)
	pele escura x outras	0,0014	14,0659 (2,6881 - 73,6018)
promo15	cabelos pretos x outros	0,0172	20,3531 (1,0453 - 396,2980)
Codificante			
cds5	pele escura x pele clara	0,0002	21,8824 (2,7033 - 177,1274)
	pele escura x pele media	0,0003	20,6275 (2,5478 - 167,0030)
	pele escura x outras	1,71 x 10⁻⁰⁵	21,2549 (4,4492 - 101,5395)
cds7	pele escura x pele clara	0,0061	23,7042 (1,2654 - 444,0331)
cds9	pele escura x pele clara	0,0017	29,2464 (1,6033 - 533,4993)
cds12	ausência x presença sardas	0,0150	21,6144 (1,1105 - 420,7012)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Tanto para a região promotora quanto para a codificante foi considerado $\alpha_{\text{corrigido}} = 0,0026$ (isto é, $0,05/19$ haplótipos).

6.5 Associações índices de melanina

Para verificar a existência de associação entre uma determinada variante e a variação no índice de melanina, foram considerados apenas os genótipos referentes às 28 variantes do gene *SLC45A2* e 15 do gene *TYR* previamente associadas a alguma das características de pigmentação (Tabelas 13 e 14, respectivamente). Os resultados de associação baseados no índice de melanina estão apresentados nas Tabelas 17 e 18. Podemos observar que para o gene *SLC45A2*, a maioria dos 28 SNPs se mostrou significativamente associada à variação dos índices de melanina, com exceção dos SNPs rs45552240 (região da testa e braço), rs113928135 e rs58784954 (regiões da testa) e rs250413 (região braço) (Tabela 17). Já para o gene *TYR*, os genótipos de todos os SNPs estão associados de forma significativa à variação dos índices de melanina tanto na testa quanto no braço, com exceção do SNP localizado na posição 88907815 (Tabela 18).

Tabela 17: Distribuição dos índices de melanina de acordo com os genótipos, considerando os 28 SNPs do gene *SLC45A2* previamente associados a alguma das características de pigmentação avaliadas (Tabela 14). R/R representam indivíduos com o genótipo homocigoto alelo referência, R/A são indivíduos com genótipo heterocigoto e A/A são indivíduos com genótipo homocigoto para o alelo alternativo.

SNPs	REF	ALT	Testa						Braço									
			Média do índice de melanina			Mann-Whitney	Kruskal-Wallis	p de R/R x R/A	p de R/R x A/A	p de R/A x A/A	Média do índice de melanina			Mann-Whitney	Kruskal-Wallis	p de R/R x R/A	p de R/R x A/A	p de R/A x A/A
			R/R	R/A	A/A	p	p				R/R	R/A	A/A	p	p			
rs45552240	T	G	1,475	1,389		0,5247				0,9973	0,7667		0,1341					
rs143719822	C	CA	1,444	1,893		<0,0001				0,9823	1,526		<0,0001					
rs35406	A	G	1,438	1,674	1,356		0,0011	<0,001	>0,05	>0,05	0,9807	1,107	0,8647		0,0032	<0,01	>0,05	>0,05
rs7380538	A	G	2,158	1,848	1,427		<0,0001	>0,05	>0,05	<0,001	1,901	1,287	0,9720		<0,0001	>0,05	>0,05	<0,001
rs35407	A	G	1,922	1,675	1,314		<0,0001	<0,05	<0,001	<0,001	1,486	1,188	0,8483		<0,0001	>0,05	<0,001	<0,001
rs57959046	C	A	1,442	1,906	2,242		<0,0001	<0,001	>0,05	>0,05	0,9827	1,414	2,224		<0,0001	<0,001	>0,05	>0,05
rs55746663	A	G	1,440	1,893	2,242		<0,0001	<0,001	>0,05	>0,05	0,9823	1,438	2,224		<0,0001	<0,001	>0,05	>0,05
rs6874998	T	C	1,459	2,047		0,0028				0,9933	1,740		0,0070					
rs6894408	C	G	1,450	1,941		<0,0001				0,9883	1,671		<0,0001					
rs250416	C	A	1,437	1,696	1,353		0,0004	<0,001	>0,05	>0,05	0,9793	1,141	0,8532		0,0018	<0,01	>0,05	>0,05
rs16891982	C	G	1,906	1,672	1,265		<0,0001	<0,05	<0,001	<0,001	1,429	1,181	0,7987		<0,0001	<0,05	<0,001	<0,001
rs35398	G	T	1,438	1,669	1,353		0,0011	<0,01	>0,05	>0,05	0,9807	1,084	0,8532		0,0046	<0,01	>0,05	>0,05
rs113928135	G	A	1,463	1,853		0,0669				0,9933	1,315		0,0414					
rs2287949	T	C		1,846	1,433		<0,0001				1,274	0,9778		<0,0001				
rs1010872	A	C	2,125	1,849	1,427		<0,0001	>0,05	>0,05	<0,001	1,745	1,299	0,9720		<0,0001	>0,05	>0,05	<0,001
rs28777	C	A	1,881	1,687	1,308		<0,0001	>0,05	<0,001	<0,001	1,439	1,213	0,8380		<0,0001	>0,05	<0,001	<0,001
rs141348470	AAAC	A	1,449	1,848	1,432		<0,0001	>0,05	>0,05	<0,001	1,054	1,287	0,9793		0,0002	>0,05	>0,05	<0,001
rs26722	C	T	1,436	1,668	1,702		0,0015	<0,01	>0,05	>0,05	0,9823	1,075	1,222		0,0377	<0,05	>0,05	>0,05
rs376784068	T	A	1,466	2,134		0,0359				0,9963	2,041		0,0105					
rs183671	T	G	1,880	1,690	1,307		<0,0001	>0,05	<0,001	<0,001	1,361	1,216	0,8467		<0,0001	>0,05	<0,001	<0,001
rs6451048	A	G		2,106	1,460		0,0058				1,828	0,9960		0,0137				
rs6451049	T	C	1,449	1,846	1,434		<0,0001	>0,05	>0,05	<0,001	1,054	1,274	0,9798		0,0005	>0,05	>0,05	<0,001
rs58784954	G	A	1,463	2,112		0,0546				0,9950	1,856		0,0328					
rs250413	G	A	1,453	1,789	1,608		0,0238	<0,05	>0,05	>0,05	0,9963	1,130	0,9355		0,2275			
rs551953158	C	T	1,466	2,269		0,0036				0,9960	2,042		0,0080					
rs73077153	C	T	1,466	2,294		0,0018				0,9960	2,128		0,0039					
rs73077154	G	C	1,447	1,950		<0,0001				0,9815	1,560		<0,0001					
rs143386685	G	A	1,463	2,048		0,0124				0,9950	1,742		0,0256					

Tabela 18: Distribuição dos índices de melanina de acordo com os genótipos, considerando os 15 SNPs do gene *TYR* previamente associados a alguma das características de pigmentação avaliadas (Tabela 13). R/R representam indivíduos com o genótipo homocigoto alelo referência, R/A são indivíduos com genótipo heterocigoto e A/A são indivíduos com genótipo homocigoto para o alelo alternativo.

SNPs	REF	ALT	Testa							Braço								
			Mediana do índice de melanina			Mann-Whitney	Kruskal-Wallis	<i>p</i> de R/R x R/A	<i>p</i> de R/R x A/A	<i>p</i> de R/A x A/A	Mediana do índice de melanina			Mann-Whitney	Kruskal-Wallis	<i>p</i> de R/R x R/A	<i>p</i> de R/R x A/A	<i>p</i> de R/A x A/A
			R/R	R/A	A/A	<i>p</i>	<i>p</i>				R/R	R/A	A/A	<i>p</i>	<i>p</i>			
rs75741664	C	T	1,459	1,996		0,0022						0,9927	1,480		0,0383			
88907815	A	G	1,475	1,238		0,1767						0,9973	0,7835		0,2463			
rs142881018	CCTGA	C	1,459	2,074		0,0003						0,9890	1,880		<0,0001			
rs61759512	G	A	1,458	1,996		0,0021						0,9933	1,480		0,0391			
rs11824466	C	T	1,455	2,112	2,044		0,0008	<0,01	>0,05	>0,05		0,9900	1,861	1,386	0,0023	<0,01	>0,05	>0,05
rs55870815	T	C	1,460	2,023		0,0245						0,9960	1,560		0,0364			
rs1042602	C	A	1,571	1,414	1,435		<0,0001	<0,001	<0,05	>0,05		1,198	0,9355	0,8942	<0,0001	<0,001	<0,01	>0,05
rs11018530	C	G	1,426	1,663	2,123		<0,0001	<0,001	<0,001	<0,05		0,9670	1,160	1,783	<0,0001	<0,01	<0,001	<0,05
rs67512950	ACTT	A	1,432	1,660	2,105		0,0001	<0,01	<0,01	>0,05		0,9763	1,157	1,586	0,0002	<0,05	<0,01	>0,05
rs12804012	G	A	1,422	1,663	2,123		<0,0001	<0,001	<0,001	<0,01		0,9520	1,157	1,797	<0,0001	<0,01	<0,001	<0,01
rs75375648	T	C	1,458	2,205		<0,0001						0,9918	1,773		0,0001			
rs113438665	T	C	1,455	2,018		0,0006						0,9900	1,430		0,0018			
rs1393350	G	A	1,558	1,420	1,255		<0,0001	<0,001	<0,01	>0,05		1,076	0,9443	0,6430	<0,0001	<0,01	<0,01	>0,05
rs1126809	G	A	1,564	1,422	1,251		<0,0001	<0,01	<0,01	>0,05		1,104	0,9420	0,6283	<0,0001	<0,001	<0,01	>0,05
rs28521275	C	T	1,424	1,459	1,712		0,0003	>0,05	<0,001	<0,01		0,9272	0,9827	1,361	<0,0001	>0,05	<0,001	<0,001

6.6 Análise de miRNAs

O gene *SLC45A2* possui três diferentes 3'UTRs de acordo com suas isoformas. Para a isoforma 3, apenas um sítio de variação foi encontrado na região 3'UTR, o rs16891982, que possui como alelo referência uma citosina e como alelo alternativo, uma guanina. Dessa forma, denominou-se o haplótipo que contém o alelo referência como "alelo C" e o haplótipo que contém o alelo alternativo como "alelo G". Os resultados da análise de busca por miRNAs com diferentes afinidades a estes haplótipos foram ordenados segundo três critérios: E (energia), S (especificidade) e ES (energia e especificidade). Os 20 miRNAs encontrados para esta região com maiores valores de ΔE , ΔS e ΔES estão na Tabela 19.

Para a isoforma 1, dois pontos de variação foram encontrados na região 3'UTR menor gerando três haplótipos diferentes ("Iso1utr1", "Iso1utr2" e "Iso1utr3"). Os haplótipos "Iso1utr2" e "Iso1utr3" foram considerados nessa análise apesar de possuírem frequências menores do que 1% na nossa população (0,59% e 0,44%, respectivamente). Usando os mesmos critérios da 3'UTR com apenas um SNP, os 20 miRNAs encontrados para esta região com maiores valores de ΔE , ΔS e ΔES estão na Tabela 20.

E por fim, para a isoforma 2, nove pontos de variação foram encontrados na região 3'UTR maior gerando sete haplótipos diferentes com frequência maior que 1% na presente amostra populacional. A análise de miRNAs usando os mesmos critérios anteriores ordenados segundo os maiores valores de ΔE , ΔS e ΔES estão presentes na Tabela 21. Com o objetivo de deixar a tabela mais apresentável esteticamente, os nomes dos haplótipos ficaram da seguinte forma: "Iso2utr1" = "utr1", "Iso2utr2" = "utr2", "Iso2utr3" = "utr3", "Iso2utr4" = "utr4", "Iso2utr5" = "utr5", "Iso2utr6" = "utr6" e "Iso2utr7" = "utr7".

Tabela 19: Relação dos 20 miRNAs que apresentaram maiores valores de ΔE , ΔS e ΔES para a região 3'UTR com apenas um SNP (rs16891982) do gene *SLC45A2*.

miRNA	Score E		
	Alelo C	Alelo G	ΔE
miR-3127-5p	0,8492	0,5347	0,3146
miR-152-5p	0,7115	0,4811	0,2305
miR-6783-3p	0,5341	0,7374	0,2033
miR-4425	0,5626	0,7656	0,2030
miR-876-3p	0,3650	0,5666	0,2016
miR-6825-5p	1,0142	0,8140	0,2002
miR-3170	0,5912	0,7847	0,1936
miR-149-5p	0,9980	0,8192	0,1788
miR-6875-3p	0,4307	0,6046	0,1739

miR-6852-5p	0,5489	0,7215	0,1726
miR-6738-3p	0,5622	0,7346	0,1724
miR-210-3p	0,5808	0,4086	0,1722
miR-1293	0,9833	0,8118	0,1716
miR-450a-2-3p	0,7385	0,9087	0,1702
miR-3182	0,5397	0,7024	0,1627
miR-6503-5p	0,5039	0,6384	0,1345
miR-4455	0,6648	0,5450	0,1198
miR-3165	0,6014	0,4816	0,1197
miR-27a-5p	0,6273	0,6994	0,0721
miR-370-3p	0,4868	0,5521	0,0654

Score S			
miRNA	Alelo C	Alelo G	ΔS
miR-450a-2-3p	1,0076	2,4901	1,4824
miR-876-3p	0,1488	0,8632	0,7144
miR-149-5p	1,2143	0,5012	0,7131
miR-152-5p	0,8676	0,1767	0,6909
miR-3182	0,4176	1,0903	0,6726
miR-1293	0,9906	0,3258	0,6648
miR-4425	0,1975	0,6262	0,4287
miR-3170	0,2008	0,5570	0,3562
miR-6825-5p	0,5606	0,2305	0,3300
miR-3127-5p	0,4887	0,1587	0,3300
miR-4455	0,4817	0,2610	0,2207
miR-6783-3p	0,1702	0,3557	0,1855
miR-6738-3p	0,1887	0,3723	0,1836
miR-3691-5p	0,4938	0,6680	0,1742
miR-6852-5p	0,1546	0,3280	0,1734
miR-6503-5p	0,1945	0,3254	0,1309
miR-3165	0,2837	0,1678	0,1160
miR-27a-5p	0,2220	0,3196	0,0977
miR-210-3p	0,2282	0,1409	0,0873
miR-6875-3p	0,1407	0,2227	0,0820

Score ES			
miRNA	Alelo C	Alelo G	ΔES
miR-450a-2-3p	2,6447	4,4776	1,8328
miR-152-5p	2,4534	1,2121	1,2412
miR-876-3p	0,9346	2,0949	1,1603
miR-149-5p	3,4347	2,2845	1,1502
miR-1293	3,1976	2,1028	1,0948
miR-3127-5p	2,3755	1,3509	1,0246
miR-3182	1,6068	2,6218	1,0149
miR-4425	1,4299	2,2924	0,8626
miR-6825-5p	2,8167	2,0036	0,8131
miR-3170	1,4957	2,2656	0,7699
miR-6783-3p	1,3054	1,9369	0,6315
miR-6738-3p	1,4131	1,9706	0,5575
miR-6852-5p	1,3511	1,8929	0,5418

miR-4455	1,9451	1,4585	0,4866
miR-210-3p	1,4844	1,0225	0,4618
miR-6875-3p	1,0692	1,5303	0,4611
miR-6503-5p	1,2888	1,7022	0,4133
miR-3165	1,5793	1,2096	0,3697
miR-3691-5p	2,0829	2,3430	0,2601
miR-27a-5p	1,5831	1,8298	0,2467

Tabela 20: Relação dos 20 miRNAs que apresentaram maiores valores de ΔE , ΔS e ΔES para os três haplótipos da região 3'UTR isoforma 1 do gene *SLC45A2*.

Score E				
miRNA	Iso1utr1	Iso1utr2	Iso1utr3	ΔE
miR-4639-3p	0,5618	0,8311	0,5536	0,8175
miR-641	0,5417	0,8175	0,5465	0,8175
miR-744-3p	0,8898	0,6416	0,8816	0,6280
miR-708-5p	0,5626	0,3378	0,5692	0,3378
miR-1185-2-3p	0,4740	0,7054	0,4792	0,7054
miR-3173-5p	0,6303	0,4159	0,6221	0,6221
miR-345-5p	0,7300	0,5204	0,7229	0,5205
miR-34a-5p	0,7488	0,5524	0,7555	0,5524
miR-143-3p	0,6463	0,4566	0,6530	0,4566
miR-423-5p	0,8370	0,6571	0,8436	0,6571
miR-204-5p	0,5172	0,3520	0,5090	0,3390
miR-1291	0,9577	0,8000	0,9506	0,7862
miR-766-5p	0,7688	0,6054	0,7755	0,6054
miR-193b-3p	0,5541	0,7206	0,5541	0,7158
miR-182-5p	0,6073	0,7734	0,6140	0,7735
miR-34b-5p	0,6956	0,5351	0,7005	0,5351
miR-4431	0,6550	0,5082	0,6479	0,4945
miR-301a-5p	0,6553	0,5084	0,6471	0,4948
miR-34c-5p	0,5801	0,4264	0,5850	0,4264
miR-378a-5p	0,6718	0,5133	0,6654	0,5134
Score S				
miRNA	Iso1utr1	Iso1utr2	Iso1utr3	ΔS
miR-539-3p	0,9955	0,9955	1,3753	0,9955
miR-514b-5p	0,3202	0,3202	0,5709	0,3202
miR-487a-3p	0,4876	0,4876	0,6780	0,4876
miR-921	0,4000	0,4000	0,5574	0,4000
miR-376b-3p	0,3068	0,3068	0,4319	0,3068
miR-889-3p	0,9378	0,9378	1,0225	0,9378
miR-561-5p	0,7570	0,7570	0,8396	0,7570
miR-34c-3p	0,3810	0,3810	0,3006	0,3810
miR-616-3p	0,3187	0,3187	0,2385	0,3187
miR-3129-3p	0,8753	0,8753	0,7957	0,8753
miR-4637	1,0755	1,0755	1,0102	1,0755
miR-146b-3p	0,6257	0,6257	0,5639	0,6257
miR-433-5p	0,5807	0,5807	0,6408	0,5807
miR-212-5p	0,6090	0,6090	0,5505	0,6090

miR-3167	0,4186	0,4186	0,3612	0,4186
miR-548b-3p	1,0398	1,0398	0,9826	1,0398
miR-301a-5p	0,9396	0,9396	0,8838	0,9396
miR-4755-5p	0,5214	0,5214	0,4664	0,5214
miR-573	0,6559	0,6559	0,7096	0,6559
miR-3133	1,2373	1,2373	1,2909	1,2373

Score ES

miRNA	Iso1utr1	Iso1utr2	Iso1utr3	Δ ES
miR-539-3p	2,2496	2,2496	2,7384	2,2496
miR-514b-5p	1,7977	1,7977	2,1842	1,7977
miR-487a-3p	1,8323	1,8323	2,1243	1,8323
miR-376b-3p	1,3638	1,3638	1,5916	1,3638
miR-921	2,0671	2,0671	2,2879	2,0671
miR-616-3p	1,9491	1,9491	1,7368	1,9491
miR-34c-3p	1,5861	1,5861	1,4494	1,5861
miR-3129-3p	2,1092	2,1092	2,0030	2,1092
miR-889-3p	2,1079	2,1079	2,2116	2,1079
miR-561-5p	2,0011	2,0011	2,0954	2,0011
miR-4637	2,4599	2,4599	2,3676	2,4599
miR-146b-3p	2,3424	2,3424	2,2588	2,3424
miR-212-5p	2,1730	2,1730	2,0931	2,1730
miR-548b-3p	2,5469	2,5469	2,4672	2,5469
miR-3167	1,4727	1,4727	1,3944	1,4727
miR-301a-5p	2,4347	2,4347	2,3573	2,4347
miR-4755-5p	1,9820	1,9820	1,9078	1,9820
miR-433-5p	2,1362	2,1362	2,2068	2,1362
miR-3661	2,9049	2,9049	2,8348	2,9049
miR-99a-5p	2,0316	2,0316	1,9626	2,0316

Tabela 21: Relação dos 20 miRNAs que apresentaram maiores valores de Δ E, Δ S e Δ ES para os sete haplótipos da região 3'UTR isoforma 2 do gene *SLC45A2*.

Score E

miRNA	utr1	utr2	utr3	utr4	utr5	utr6	utr7	Δ E
miR-197-3p	0,5944	1,3321	1,3321	1,3321	1,3321	0,5944	1,3299	0,7377
miR-766-3p	1,3609	0,6859	0,6859	1,3609	0,6859	1,3609	1,3587	0,6750
miR-664b-3p	0,6664	0,6585	0,6585	0,6685	1,2007	0,6664	0,6685	0,5421
miR-203a-5p	0,4119	0,9445	0,9445	0,4927	0,9417	0,4119	0,4927	0,5326
miR-6515-3p	0,2973	0,2887	0,2887	0,3232	0,7628	0,2973	0,3232	0,4741
miR-671-3p	0,4813	0,4748	0,4748	0,5735	0,9193	0,4813	0,5735	0,4445
miR-6738-3p	0,5383	0,5383	0,5383	0,9650	0,9650	0,5383	0,9644	0,4266
miR-146a-3p	0,7239	0,3365	0,7270	0,7594	0,7280	0,7239	0,7594	0,4229
miR-1180-5p	0,5754	0,5627	0,5627	0,5668	0,9823	0,5754	0,5668	0,4195
miR-92b-3p	0,9259	0,5119	0,9259	0,9259	0,9259	0,9259	0,9233	0,4141
miR-92b-3p	0,9259	0,5119	0,9259	0,9259	0,9259	0,9259	0,9233	0,4141
miR-4709-5p	0,8994	0,5720	0,5720	0,5333	0,5937	0,8994	0,5333	0,3661
miR-5699-3p	0,6540	0,6222	0,6222	0,9025	0,7000	0,6540	0,9762	0,3540

miR-3129-5p	0,5059	0,7020	0,7020	0,7020	0,5397	0,5059	0,8459	0,3401
miR-6514-3p	0,8674	0,5478	0,8674	0,8674	0,8674	0,8674	0,8837	0,3359
miR-6780a-3p	0,5261	0,5172	0,5172	0,5265	0,8370	0,5261	0,5265	0,3198
miR-505-5p	0,8785	0,8785	0,8785	0,5599	0,8689	0,8646	0,5599	0,3186
miR-506-5p	0,5282	0,5215	0,5215	0,8360	0,5236	0,5282	0,8363	0,3148
miR-6775-3p	0,7672	0,7682	0,7682	0,8020	1,0784	0,7672	0,8020	0,3112
miR-3151-5p	1,1575	1,1575	1,1575	0,8463	0,9060	1,1544	0,8463	0,3112

Score S

miRNA	utr1	utr2	utr3	utr4	utr5	utr6	utr7	ΔS
miR-197-3p	0,3211	2,8996	2,8996	2,8996	2,8996	0,3211	2,8979	2,5785
miR-766-3p	2,8105	0,2914	0,2914	2,8105	0,2914	2,8105	2,8082	2,5191
miR-664b-3p	0,5397	0,5363	0,5363	0,5412	2,8822	0,5397	0,5412	2,3459
miR-203a-5p	0,1365	2,4674	2,4674	0,1953	2,4325	0,1365	0,1953	2,3309
miR-4709-5p	2,0330	0,3121	0,3121	0,2431	0,3136	2,0330	0,2431	1,7900
miR-5699-3p	0,3292	0,4196	0,4196	1,6169	0,4296	0,3292	2,0833	1,7540
miR-506-5p	0,2169	0,2115	0,2115	1,8783	0,2135	0,2169	1,8783	1,6667
miR-299-3p	0,3034	1,5288	1,5288	1,7892	0,2243	0,3034	1,7892	1,5648
miR-659-5p	0,6584	0,6584	0,6584	0,6584	2,1277	0,5990	0,5933	1,5344
miR-4782-5p	2,3819	2,3819	2,3819	2,3677	0,8795	2,3819	2,3677	1,5024
miR-146a-3p	1,4090	0,1356	1,4160	1,5825	1,4233	1,4090	1,5825	1,4470
miR-1180-5p	0,2499	0,2437	0,2437	0,2449	1,6286	0,2499	0,2449	1,3849
miR-301b-3p	0,2907	0,2907	0,2907	1,5836	0,4424	0,3155	1,5836	1,2928
miR-3151-5p	1,5721	1,5721	1,5721	0,3041	0,5497	1,5509	0,3041	1,2680
miR-378a-5p	0,5859	0,5859	0,5859	0,5859	1,8443	0,6445	0,6325	1,2584
miR-3613-5p	0,2242	1,4636	0,2242	0,2242	0,2242	0,2242	1,4636	1,2394
miR-3129-5p	0,2193	0,4927	0,4927	0,4927	0,2017	0,2193	1,4406	1,2389
miR-146a-3p	0,4367	1,4160	0,4367	0,4367	0,7510	0,4367	0,1896	1,2263
miR-493-3p	1,8986	1,8986	1,8986	1,8986	2,0990	1,8757	0,9404	1,1586
miR-642b-3p	2,0395	2,0395	2,0395	2,0395	1,9653	0,8952	1,3575	1,1443

Score ES

miRNA	utr1	utr2	utr3	utr4	utr5	utr6	utr7	ΔES
miR-197-3p	1,3928	5,3356	5,3356	5,3356	5,3356	1,3928	5,3329	3,9427
miR-766-3p	5,2955	1,5341	1,5341	5,2955	1,5341	5,2955	5,2922	3,7615
miR-664b-3p	1,7486	1,7305	1,7305	1,7540	5,0725	1,7486	1,7547	3,3420
miR-203a-5p	0,8781	4,1903	4,1903	1,0865	4,1502	0,8781	1,0874	3,3123
miR-4709-5p	3,6730	1,3457	1,3457	1,2096	1,3876	3,6730	1,2105	2,4634
miR-5699-3p	1,5169	1,5358	1,5358	3,2565	1,7024	1,5169	3,8613	2,3444
miR-506-5p	1,1729	1,1552	1,1552	3,4038	1,1611	1,1729	3,4064	2,2513
miR-146a-3p	2,7262	0,7410	2,7389	2,9656	2,7482	2,7262	2,9666	2,2255
miR-1180-5p	1,2917	1,2620	1,2620	1,2707	3,4160	1,2917	1,2712	2,1540
miR-299-3p	1,3658	2,9097	2,9097	3,2894	1,1770	1,3658	3,2906	2,1136
miR-148a-5p	0,7198	2,6988	0,7198	0,7198	0,7198	0,7198	0,7201	1,9789
miR-659-5p	1,9239	1,9239	1,9239	1,9239	3,8026	1,8339	1,8257	1,9769
miR-3129-5p	1,1265	1,7699	1,7699	1,7699	1,1815	1,1265	2,9816	1,8551
miR-3151-5p	3,6858	3,6858	3,6858	1,8432	2,2012	3,6589	1,8453	1,8426
miR-6515-3p	0,6890	0,6644	0,6644	0,7749	2,4830	0,6890	0,7758	1,8187
miR-6738-3p	1,1588	1,1589	1,1589	2,9518	2,9518	1,1588	2,9525	1,7937

miR-146a-3p	1,4546	2,7389	1,4546	1,4546	1,9447	1,4546	0,9518	1,7871
miR-3613-5p	0,8213	2,6000	0,8213	0,8213	0,8213	0,8213	2,6000	1,7787
miR-4782-5p	3,7884	3,7874	3,7874	3,7671	2,0126	3,7884	3,7686	1,7758
miR-301b-3p	1,0811	1,0811	1,0811	2,8259	1,2884	1,1182	2,8272	1,7460

7. DISCUSSÃO

7.1 Aspectos populacionais e desvios do equilíbrio de Hardy-Weinberg

Resultados apresentados em um estudo realizado no Brasil sobre a estimativa de ancestralidade em indivíduos pertencentes a diferentes categorias de pigmentação e de várias regiões do país mostraram uma grande heterogeneidade dentro e entre as categorias da pigmentação e dentro e entre as regiões geográficas (Pena, Di Pietro et al. 2011). Por exemplo, os indivíduos que se auto-classificaram como “não-brancos” no Norte, Nordeste e Sudeste do Brasil, apesar de apresentaram ancestralidade europeia predominante, possuem diferentes proporções de ancestralidade africana e ameríndia. A ancestralidade europeia é ainda maior na categoria dos indivíduos que se auto-classificaram como “brancos”, variando de 66,8% na Bahia a 85,5% no Rio Grande do Sul e 86,1% no Rio de Janeiro. Este mesmo estudo aborda que a percepção de cor considera outros fatores como o formato do rosto e o tipo de cabelo, além do tom de pele, na determinação da auto-classificação. Apesar disso, de um modo geral, indivíduos com pele mais clara tendem a ter maior proporção de ancestralidade europeia e indivíduos com pele mais escura, maior proporção de ancestralidade africana. Essa tendência foi observada em estudo realizado em Brasília sobre a relação da cor da pele e a ancestralidade dos indivíduos (Leite, Fonseca et al. 2011). Em média a contribuição europeia dos indivíduos autodeclarados “brancos”, “pardos” e “negros” foi de 72%, 68% e 63%, respectivamente. E, alternativamente, a média de contribuição africana nos mesmos grupos de “brancos”, “pardos” e “negros” foi de 19%, 21% e 27%, respectivamente.

No presente estudo, dos 204 indivíduos coletados aleatoriamente, a maioria (107 indivíduos) foi classificado com pele moderadamente pigmentada (peles tipo III e IV na escala de Fitzpatrick), seguido por 72 indivíduos de pele clara (peles tipo I e II) e a categoria menos representada foi a de pele escura (peles tipo V e VI, 25 indivíduos). Quando comparamos esses dados com os dados baseados nos critérios adotados pelo IBGE, encontramos: 133 indivíduos classificados pelo grupo de pesquisa como brancos (dentre eles, 70 indivíduos tinham pele clara e 59, pele tipo III); 3 amarelos (dois de pele III); 53 pardos, dentre eles, a grande maioria de pele média (cinco com

pele III e 38 com pele IV) e 15 pretos (todos com pele escura). Se levarmos em consideração que a variação na pigmentação da pele é uma característica contínua e que adotar categorias é uma decisão de certa forma arbitrária, a distribuição dos tipos de pele ao longo das categorias baseadas nos critérios do IBGE se mostrou pertinente, havendo um aumento gradual dos tipos de pele mais claros para os mais escuros dos indivíduos brancos para os pretos.

Ainda em relação a amostra coletada aleatoriamente, quatro (rs35407, rs16891982, rs28777 e rs183671) dos 58 *loci* avaliados para o gene *SLC45A2* e um (rs11824466) dos 42 *loci* avaliados para o gene *TYR* apresentaram desvios com relação ao equilíbrio de Hardy-Weinberg. Todos os *loci* que apresentaram desvios possuem excesso de homozigotos em relação ao equilíbrio em ambos os genes. Para o gene *SLC45A2*, era estatisticamente esperado que 2,9 casos apresentassem desvio em 58 análises já que aproximadamente 5% das análises resultam em desvios casuais ao se adotar o nível de significância de 0,05; entretanto, os valores observado e esperado não diferem significativamente ($\chi^2 = 0,4174$; $p = 0,5182$). Contudo, uma vez que estas variantes estão ligadas, um possível fator que pode ter ocasionado esse desvio é a ocorrência de mistura inter-étnica, associada à subestrutura populacional. A alta miscigenação da população de Ribeirão Preto associada à ocorrência de casamentos preferenciais no que se refere a ancestralidade e pigmentação da pele, pode resultar em estratificação populacional (efeito Wallund). Apesar de que evidências de seleção positiva agindo sob o SNP rs16891982, favorecendo fenótipos de pigmentação clara em europeus nos últimos 5000 anos, tenham sido encontradas (Wilde, Timpson et al. 2014), é improvável que sua intensidade seja suficiente para resultar em desvios do equilíbrio de Hardy-Weinberg em uma amostra populacional brasileira. Para o *locus* do gene *TYR*, todavia, é mais provável que esse único desvio seja devido tão somente ao acaso, já que era esperado uma ocorrência de 2,1 casos de desvio em 42 análises.

7.2 Associações encontradas

Estudos genéticos sobre a variação fenotípica normal em humanos, ratos e peixes identificaram muitos genes que regulam a pigmentação (Wasmeier, Hume et al. 2008). Essa regulação é um processo complexo que pode ser controlado em diferentes níveis (Pillaiyar, Manickam et al. 2017). Em mamíferos, complexos enzimáticos constituídos por produtos de genes específicos de melanócitos agem coordenando de forma rigorosa a via da melanogênese e a melanina é sintetizada por uma série de

interações catalisadas por esses complexos (D'Mello, Finlay et al. 2016), que incluem os produtos dos genes *SLC45A2* e *TYR*.

Somente recentemente mais informações sobre a estrutura e função desses genes (*SLC45A2* e *TYR*) foram publicadas (Bin, Bhin et al. 2015, Kondo, Namiki et al. 2015, Lai, Wichers et al. 2018) e, embora associações entre seus alelos/haplótipos e fenótipos venham sendo publicadas há algum tempo, elas estão concentradas em populações mais homogêneas e variantes polimórficas isoladas. Neste sentido, as associações identificadas nas diferentes regiões em ambos os genes no presente trabalho podem contribuir para aumentar as informações sobre variantes pouco frequentes e melhorar a compreensão sobre a diversidade desses genes em populações miscigenadas.

7.2.1 Gene *SLC45A2*

7.2.1.1 REGIÃO PROMOTORA

Quatro variantes da região promotora foram associadas com fenótipos de pigmentação. Três delas (rs143386685, rs73077153 e rs551953158) possuem alelo alternativo raro e não são consideradas polimórficas na população coletada aleatoriamente (Tabela 11). Já o SNP rs73077154 apresentou seu alelo alternativo associado a fenótipos escuros, assim como todos os outros três alelos alternativos. Os alelos rs73077153*T ($p = 0,0269$, OR = 25,72), rs551953158*T ($p = 0,0269$, OR = 25,72) e rs73077154*C ($p = 2,34 \times 10^{-06}$, OR = 18,78) se mostraram associados com pele escura e o alelo rs143386685*A ($p = 0,0160$, OR = 21,00) com cabelos pretos (Tabela 13).

O Projeto 1000 Genomas, que foi realizado entre 2008 e 2015, é responsável pelo maior catálogo público de dados genotípicos e de variantes genéticas humanas em uma escala global (Genomes Project, Auton et al. 2015). Em sua fase final, foram reconstruídos os genomas de 2.504 indivíduos de 26 populações (divididas em 5 grandes grupos populacionais: Africanos - AFR, Populações miscigenadas americanas - AMR, Leste-asiáticos - EAS, Europeus - EUR e Sul-asiáticos - SAS). Podemos observar na Tabela 22 que os alelos alternativos dos SNPs rs143386685, rs73077153 e rs551953158 são raros em todas as populações mundiais, exceto em africanos em que os alelos rs143386685*A e rs73077153*T apresentaram frequências de 1,97% e 3,30%, respectivamente, concordando com suas associações a fenótipos escuros. Estas informações são compatíveis com os resultados das análises de índice de melanina em que os genótipos contendo os alelos alternativos rs143386685*G/A e

rs73077153*C/T foram associados no presente estudo a um aumento significativo do índice de melanina em comparação aos grupos com genótipos homocigoto referência rs143386685*G/G e rs73077153*C/C (Tabela 17).

Tabela 22: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados na região promotora do gene *SLC45A2*, nas populações mundiais e na população do presente estudo.

Polimorfismo	Alelo REF ALT	BRA (n=200)	EAS (n=504)	EUR (n=503)	AFR (n=661)	AMR (n=347)	SAS (n=489)
rs143386685	G	0,9950	1,0000	1,0000	0,9803	0,9971	1,0000
	A	0,0050	-	-	0,0197	0,0029	-
rs73077154	G	0,9725	1,0000	1,0000	0,7746	0,9798	1,0000
	C	0,0275	-	-	0,2254	0,0202	-
rs73077153	C	0,9975	1,0000	1,0000	0,9667	0,9986	1,0000
	T	0,0025	-	-	0,0330	0,0014	-
rs551953158	C	1,0000	1,0000	1,0000	0,9985	1,0000	1,0000
	T	-	-	-	0,0015	-	-

Entre os haplótipos encontrados associados com fenótipos de pigmentação nesta região, o haplótipo promo3, apesar de associado com cabelos pretos ($p = 0,0001$, OR = 15,83), possui todos os alelos referência desses quatro SNPs. Já os haplótipos promo15 e promo17, também associados com fenótipos escuros [cabelos pretos ($p = 0,0174$, OR = 20,24) e pele escura ($p = 0,0245$, OR = 27,24), respectivamente], possuem um (rs143386685*A) e dois (rs73077154*C, rs73077153*T) alelos alternativos, respectivamente, embora as associações encontradas para eles sejam mais fracas que as encontradas para o haplótipo promo3 (Tabela 15). Sabe-se que a regulação da expressão genica é muito complexa e envolve diversos mecanismos, incluindo a ligação de fatores de transcrição a sequências regulatórias no DNA (Talebzadeh e Zare-Mirakabad 2014). Utilizando a ferramenta HaploReg v4.1 (<http://archive.broadinstitute.org/mammals/haploreg/haploreg.php>) para busca de SNPs em regiões não codificadoras do genoma, o SNP rs551953158 não foi encontrado pela ferramenta e os outros três SNPs (rs143386685, rs73077154 e rs73077153) foram encontrados em região sobreposta com parte da região 3'UTR do gene *AMACR*. Este gene codifica a enzima alfa metilacil CoA racemase, que é encontrada em mitocôndrias e peroxissomos e está envolvida na biossíntese do ácido biliar e no metabolismo de lipídios (Li, Fang et al. 2014), sendo que suas mutações já foram relacionadas com câncer de próstata (Chen, Wang et al. 2015). Efeitos de variações em *motifs* regulatórios podem ser preditos pela diferença entre os valores de *LOD* (*Log of the Odds*) entre os alelos alternativo e referência ($LOD_{alt} - LOD_{ref}$). Valores negativos indicam alta afinidade do fator de transcrição com o alelo referência,

enquanto que valores positivos sugerem alta afinidade com o alelo alternativo (Zhang, Zhou et al. 2016). No caso dos SNPs (rs143386685, rs73077154 e rs73077153) encontrados pela ferramenta HaploReg, apenas dois deles se encontram em sítios de ligação de fatores de transcrição. Foram encontrados quatro fatores de transcrição para o SNP rs143386685 (COMP1, E4F1, Myc_disc5, Nrf-2_2) e três para o SNP rs73077153 (Foxc1_1, Hoxa5_2, Pax-8_2). No caso do SNP rs143386685, o fator de transcrição E4F1 se liga com uma afinidade 12 vezes maior ao alelo referência (rs143386685*G) do que com o alelo alternativo (rs143386685*A). Já os outros fatores se ligam com maior afinidade ao alelo alternativo, sendo que o que se liga com maior afinidade foi o fator Nrf-2_2 (11,9). É importante mencionar que estes dois fatores (E4F1 e Nrf-2_2) são expressos na pele (<https://www.proteinatlas.org/>). Sabendo que a função do fator Nrf-2_2 é proteger as células epiteliais contra a ação de espécies reativas de oxigênio, que a radiação UV leva a um aumento destas e que isto pode levar ao estresse oxidativo causando envelhecimento, morte celular e câncer (Gegotek e Skrzydlewska 2015), podemos entender a relação de maior afinidade desse fator com o alelo associado a fenótipos mais escuros. Já o fator E4F1 é um repressor transcricional (<http://www.uniprot.org/uniprot/Q66K89>) e sua maior afinidade com o alelo referência é compatível com a determinação de fenótipos de pigmentação mais claros. Em relação ao SNP rs73077153, os fatores Hoxa5_2 (10,3) e Foxc1_1 (0,7) se ligam com maior afinidade ao alelo referência (rs73077153*C) e o fator Pax-8_2 (0,8) ao alelo alternativo (rs73077153*T). Apenas os fatores Hoxa5_2 e Foxc1_1 se expressam na pele, porém a expressão em melanócitos é baixa ou inexistente, respectivamente (<https://www.proteinatlas.org/>). A não ser que tais fatores sejam inibitórios, a associação encontrada do alelo rs73077153*C com pigmentação clara é contraditória. Entretanto, a baixa expressão de tais fatores em melanócitos indica que sua influência na variação fenotípica encontrada é pequena.

7.2.1.2 REGIÃO CODIFICANTE E 3'UTR

Foram encontradas 24 variantes associadas com fenótipos nessas regiões dependendo da isoforma considerada. Treze variantes estão em íntrons (rs250413, rs58784954, rs6451049, rs6451048, rs183671, rs376784068, rs141348470, rs28777, rs1010872, rs113928135, rs35398, rs250416, rs146922924) e uma variante (rs2287949) em éxon nas três isoformas. O SNP rs45552240 encontra-se apenas na região 3'UTR da isoforma1. Dentre as nove variantes restantes, sete (rs6894408, rs6874998, rs55746663, rs57959046, rs35407, rs7380538 e rs35406) estão localizadas no íntron seis da isoforma 1 ou na região 3'UTR da isoforma 2; o SNP rs26722 se localiza no éxon três das isoformas 1 e 2 ou no íntron dois da isoforma 3 e

o SNP rs16891982 está no éxon cinco das isoformas 1 e 2 ou na região 3'UTR da isoforma 3.

Podemos observar que a maior parte da diversidade do gene *SLC45A2* encontra-se na região intrônica independente da isoforma. Como já mencionado, íntrons podem desempenhar funções de regulação da expressão gênica ao afetarem eventos de *splicing* e/ou a estabilidade do RNAm através de sua interação com proteínas de regulação e outros RNAs (por exemplo, microRNAs) (Wang, Feng et al. 2007, Mishra e Thakran 2018), e sendo assim, polimorfismos nessa região pode influenciar a variação na pigmentação. Não se pode descartar a possibilidade de que tais associações em polimorfismos intrônicos sejam decorrentes do elevado desequilíbrio de ligação com outras variantes. Encontramos quatro SNPs não polimórficos associados a algum fenótipo nessa região. Os alelos desses SNPs encontrados em menor frequência na nossa amostra se mostraram completamente ausentes ou monomórficos em todas as populações mundiais, exceto na população africana em que o rs58784954*A teve 6,58% de frequência, seguido dos alelos rs113928135*A com 3,18% e rs6451048*A com 2,8% (Tabela 23). Em conformidade com esses dados, o alelo menos frequente de todos eles está associado com fenótipos escuros na nossa população: os alelos rs58784954*A ($p = 0,0154$, OR = 15,60) e rs376784068*A ($p = 0,0272$, OR = 25,54) estão associados a pele escura e o alelo rs113928135*A a cabelos pretos ($p = 0,016$, OR = 12,07), enquanto que o rs6451048*A se mostrou associado a pele escura ($p = 0,0007$, OR = 46,82) e cabelos pretos ($p = 0,0041$, OR = 26,95) (Tabela 13). De forma complementar, os resultados dos índices de melanina mostram que os genótipos contendo estes alelos rs58784954*G/A, rs376784068*T/A, rs113928135*G/A, rs6451048*A/G foram associados a um aumento significativo do índice de melanina em comparação aos grupos com os genótipos rs58784954*G/G, rs376784068*T/T, rs113928135*G/G e rs6451048*G/G (Tabela 17). A ferramenta HaploReg encontrou apenas três dos SNPs mencionados (rs58784954, rs113928135 e rs6451048) presentes em sítios de ligação de fatores de transcrição. Entretanto, de todos os fatores identificados, somente Pou5f1 (maior afinidade pelo alelo rs113928135*A), Hoxa4 (maior afinidade pelo alelo rs6451048*G) e Evi-1 (maior afinidade pelo alelo rs6451048*A) se expressam na pele (<https://www.proteinatlas.org/>) e poderiam estar influenciando a variação fenotípica.

Tabela 23: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados nas regiões codificante e 3'UTR do gene *SLC45A2*, nas populações mundiais e na população do presente estudo.

Polimorfismo	Alelo	BRA (n=200)	EAS (n=504)	EUR (n=503)	AFR (n=661)	AMR (n=347)	SAS (n=489)
	REF ALT						

rs250413	G	0,9625	0,7718	0,9861	0,9380	0,9063	0,9550
	A	0,0350	0,2282	0,0139	0,0620	0,0937	0,0450
rs58784954	G	0,9925	1,0000	1,0000	0,9342	0,9971	1,0000
	A	0,0075	-	-	0,0658	0,0029	-
rs6451049	T	0,0650	0,2183	0,0139	0,1286	0,1081	0,4151
	C	0,9250	0,7817	0,9861	0,8714	0,8918	0,5849
rs6451048	A	0,0025	-	-	0,0280	-	-
	G	0,9925	1,0000	1,0000	0,9720	1,0000	1,0000
rs183671	T	0,3325	0,9086	0,0467	0,7897	0,5072	0,7710
	G	0,6625	0,0913	0,9533	0,2103	0,4928	0,2290
rs376784068	T	0,9925	1,0000	1,0000	0,9977	1,0000	1,0000
	A	0,0025	-	-	0,0023	-	-
rs26722	C	0,8800	0,6090	0,9761	0,9561	0,6916	0,7935
	T	0,1200	0,3909	0,0239	0,0439	0,3084	0,2065
rs141348470	AAAC	0,0700	0,2183	0,0159	0,1369	0,1110	0,4223
	A	0,9275	0,7817	0,9841	0,8631	0,8890	0,5777
rs28777	C	0,3250	0,8829	0,0437	0,8048	0,5072	0,7771
	A	0,6725	0,1171	0,9563	0,1952	0,4928	0,2229
rs1010872	A	0,0750	0,1796	0,0109	0,1641	0,1023	0,4213
	C	0,9250	0,8204	0,9890	0,8359	0,8977	0,5787
rs2287949	T	0,0650	0,1786	0,0109	0,1248	0,0965	0,4223
	C	0,9350	0,8214	0,9891	0,8752	0,9035	0,5777
rs113928135	G	0,9900	1,0000	1,0000	0,9682	0,9942	1,0000
	A	0,0100	-	-	0,0318	0,0058	-
rs35398	G	0,8950	0,5665	0,9761	0,9660	0,6787	0,7935
	T	0,1050	0,4335	0,0239	0,0340	0,3213	0,2065
rs16891982	C	0,3925	0,9940	0,0616	0,9644	0,5360	0,9407
	G	0,6075	0,0060	0,9383	0,0356	0,4639	0,0593
rs250416	C	0,8850	0,5675	0,9761	0,9395	0,6787	0,7935
	A	0,1100	0,4325	0,0239	0,0605	0,3213	0,2065
rs6894408	C	0,9775	1,0000	0,9990	0,9221	0,9928	1,0000
	G	0,0225	-	0,0010	0,0779	0,0072	-
rs6874998	T	0,9900	1,0000	1,0000	0,9728	0,9986	1,0000
	C	0,0100	-	-	0,0272	0,0014	-
rs55746663	A	0,9675	1,0000	1,0000	0,8389	0,9928	1,0000
	G	0,0325	-	-	0,1611	0,0072	-
rs57959046	C	0,9675	1,0000	1,0000	0,8389	0,9928	1,0000
	A	0,0325	-	-	0,1611	0,0072	-
rs35407	A	0,2975	0,8819	0,0427	0,7080	0,4885	0,7751
	G	0,7025	0,1181	0,9573	0,2920	0,5115	0,2249
rs7380538	A	0,0825	0,1776	0,0109	0,1982	0,0994	0,4223
	G	0,9175	0,8224	0,9891	0,8018	0,9006	0,5777
rs35406	A	0,8950	0,5496	0,9751	0,9667	0,6816	0,7935
	G	0,1050	0,4504	0,0249	0,0333	0,3184	0,2065
rs146922924	C	0,9700	0,9960	0,9970	0,8737	0,9784	0,9939
	CA	0,0275	0,0040	0,0030	0,1263	0,0216	0,0061
rs45552240	T	0,9975	1,0000	0,9990	0,9992	0,9986	0,9571
	G	0,0025	-	0,0010	0,0008	0,0014	0,0429

Em relação as variantes polimórficas localizadas na região intrônica, encontramos fortes associações do alelo rs28777*A com cabelos loiros ($p = 2,83 \times 10^{-24}$, OR = 50,07) e pele clara ($p = 1,01 \times 10^{-26}$, OR = 14,88). Este SNP pertence ao conjunto de SNPs utilizados pelo sistema HIrisPlex para predição das cores de olhos e cabelos

(Walsh, Liu et al. 2013) e de fato resultou na mais forte associação envolvendo regiões intrônicas. Seu alelo menos frequente na nossa população (rs28777*C – 32,5%) e que mostrou associação com os fenótipos escuros é encontrado em baixa frequência em europeus (4,37%), frequência intermediária em populações miscigenadas (AMR – 50,72%) e alta frequência em africanos (80,48%) e asiáticos (EAS – 88,29%, SAS – 77,71%) (Tabela 23). Outro SNP que apresentou fortes associações com a pigmentação na nossa amostra foi o rs183671. Seu alelo menos frequente no presente estudo (rs183671*T – 33,25%) se mostrou associado a cabelos pretos ($p = 2,39 \times 10^{-22}$, OR = 29,96) e pele escura ($p = 6,35 \times 10^{-25}$, OR = 12,88), o que está de acordo com sua frequência encontrada em populações com características de pigmentação mais escuras (AFR – 78,97%, EAS – 90,86% e SAS – 77,10%) e sua baixa frequência em europeus (4,67%). Como esperado, a distribuição das frequências desse SNP na população brasileira (rs183671*G – 66,25%) é mais semelhante às encontradas para populações miscigenadas [AMR (T – 50,72% e G – 49,28%)] do que para as outras populações mais homogêneas (Tabela 23). A ferramenta HaploReg encontrou sítios de ligação de fatores de transcrição envolvendo ambos SNPs (rs28777 e rs183671). Os fatores Nanog_disc3 e Sin3Ak-20_disc6 possuem maior afinidade pelo alelo rs28777*C, mas apenas o primeiro é expresso na pele e tem a função de ativador ou repressor transcricional (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=NANOG&keywords=Nanog>), o que é compatível com as associações encontradas para este SNP. Do mesmo modo, apenas um fator (p300_known1) se liga ao sítio envolvendo o SNP rs183671 (maior afinidade para o alelo T), sendo expresso na pele e com função de regulação da transcrição através do remodelamento da cromatina (https://www.genecards.org/cgi-bin/carddisp.pl?gene=EP300#protein_expression).

Quando levamos em consideração esses dois SNPs (rs28777 e rs183671) e as associações encontradas para os haplótipos inferidos para região codificante da isoforma 3, podemos perceber que o único haplótipo associado com fenótipos de pigmentação claros [“Iso3cds1”, pele clara ($p = 2,82 \times 10^{-25}$, OR = 17,22) e cabelos loiros ($p = 3,36 \times 10^{-23}$, OR = 50,94), Tabela 15] possui como diferença em relação aos haplótipos associados com pigmentação escura (“Iso3cds2”, “Iso3cds3”, “Iso3cds4”, “Iso3cds5”, “Iso3cds7”, “Iso3cds11” e “Iso3cds13”) os alelos rs28777*A e rs183671*G (Tabela 6). Estes SNPs estão em forte desequilíbrio de ligação entre si [LOD (*Log of the Odds*) ≥ 2 e $D' = 0,98$] e ambos com o SNP rs16891982 (éxon 5) (FIGURA 8), dessa forma, todas as associações referentes a eles podem ser devido a influência desse desequilíbrio.

A variante (rs2287949) localizada em éxon presente nas três isoformas é uma mutação sinônima (T329T). Polimorfismos localizados em éxons podem causar

variação fenotípica por diversas razões, dentre elas, a proteína mutada pode perder a função ou a união dos éxons no mRNA pode ser comprometida gerando uma isoforma diferente (Hull, Campino et al. 2007). Como a mutação no polimorfismo citado é sinônima e, portanto, não causa uma mudança de aminoácido, é possível que ela afete a união dos éxons na formação do mRNA já que se localiza próxima a um sítio de *splicing* (<http://exac.broadinstitute.org/variant/5-33954511-T-C>). Além disso, Shen e colaboradores (Shen, Basilion et al. 1999), observaram diferenças na estrutura secundária de mRNAs associadas com mutações sinônimas em SNPs em regiões codificadoras. Eles argumentam que essas diferenças podem levar a variação nas funções biológicas como resultado da interação diferencial desses mRNAs com outros componentes celulares. Posto isso, podemos observar que o alelo menos frequente (rs2287949*T – 6,5%, Tabela 23) desse SNP na nossa população está associado de forma significativa com pele escura ($p = 3,33 \times 10^{-08}$, OR = 12,30) e cabelos pretos ($p = 0,0001$, OR = 17,27), o que é congruente com sua maior frequência nas populações africana (12,48%) e asiática (EAS – 17,86% e SAS – 42,23%) e com os resultados de índice de melanina (Tabela 17): o genótipo rs2287949*T/T apresentou aumento significativo do índice de melanina em comparação ao grupo com o genótipo rs2287949*T/C.

Outros dois polimorfismos localizados em éxons (no caso das isoformas 1 e 2) se mostraram associados com a pigmentação. O SNP rs26722 é uma mutação não sinônima e causa a substituição do aminoácido lisina para glutamato. Seu alelo menos frequente (rs26722*T – 12%) na população estudada é encontrado em maiores frequências em populações asiáticas (EAS – 39,09% e SAS – 20,65%) e miscigenadas (AMR – 30,84%) (Tabela 23). No presente estudo esse mesmo alelo se mostrou associado a cabelos pretos ($p = 2,21 \times 10^{-05}$, OR = 18,86), porém, apesar de resultar em substituição de aminoácido, é importante mencionar que este SNP está em forte desequilíbrio de ligação com o SNP rs16891982 [$LOD \geq 2$ e $D' = 0,94$] (FIGURA 8) e esta associação pode não estar relacionada com a determinação da pigmentação e sim estar simplesmente refletindo o desequilíbrio entre eles. Em contrapartida, as associações haplotípicas encontradas em trabalho anterior, oriundo da dissertação de mestrado que inspirou o presente estudo, indicam que a mutação no SNP rs26722 modifica as associações no SNP rs16891982. Foi observado que para as associações significativas com olhos e pele clara que possuíam o alelo rs16891982*G (374Phe), o alelo rs26722*C fortalecia a magnitude das associações e, similarmente, para as associações do alelo rs16891982*C com fenótipos escuros, a presença do alelo rs26722*T também aumentava a magnitude das associações (Fracasso, de Andrade et al. 2017).

O SNP rs16891982 está localizado no éxon 5 (isoformas 1 e 2) e também determina uma mutação não sinônima, causando a substituição do aminoácido leucina para fenilalanina. Este é o SNP mais estudado do gene *SLC45A2*, seus alelos e genótipos já se mostraram associados com a variação normal nas características de pigmentação em vários trabalhos em diferentes populações (Graf, Hodgson et al. 2005, Yuasa, Umetsu et al. 2006, Branicki, Brudnik et al. 2008, Kayser 2015, Fracasso, de Andrade et al. 2017). Tal SNP faz parte do sistema HirisPlex de predição de cores de olhos e cabelos (Walsh, Liu et al. 2013) e já foi associado a risco de desenvolver melanoma (Ibarrola-Villava, Hu et al. 2012, Lopez, Garcia et al. 2014). Em nosso trabalho anterior, seu alelo derivado (rs16891982*G/374Phe) apresentou fortes associações com fenótipos claros: cabelos loiros ($p = 2,32 \times 10^{-10}$, OR = 11,46), pele clara ($p = 2,13 \times 10^{-21}$, OR = 5,88) e olhos azuis ($p = 1,45 \times 10^{-05}$, OR = 6,74) (Fracasso, de Andrade et al. 2017). Este mesmo alelo foi encontrado em maior frequência na população do atual trabalho (60,75%, Tabela 23) e também se mostrou ainda mais fortemente associado com fenótipos claros (Tabela 13): cabelos loiros ($p = 1,59 \times 10^{-26}$, OR = 31,72), pele clara ($p = 9,54 \times 10^{-35}$, OR = 22,66) e olhos azuis ($p = 1,29 \times 10^{-36}$, OR = 7,35), sendo esta última associação não tabelada em virtude de apresentar OR inferior a 10. A diferença na intensidade entre as associações previamente reportadas e as aqui obtidas provavelmente refletem a aleatoriedade do processo de amostragem e o maior tamanho amostral do presente estudo. Este alelo (rs16891982*G/374Phe) é encontrado em alta frequência apenas na população europeia (93,83%, Tabela 23).

Quando observamos os haplótipos inferidos para as regiões codificantes das isoformas 1 e 2 que possuem esse SNP e as associações encontradas, podemos notar que os únicos haplótipos associados com fenótipos de pigmentação claros para ambas isoformas ["Iso2cds1", pele clara ($p = 3,65 \times 10^{-29}$, OR = 22,63) e cabelos loiros ($p = 6,34 \times 10^{-23}$, OR = 26,19) e "Iso1cds1", pele clara ($p = 1,08 \times 10^{-24}$, OR = 17,38) e cabelos loiros ($p = 1,94 \times 10^{-22}$, OR = 22,93), Tabela 15] possuem como diferença em relação a todos os outros haplótipos associados com pigmentação escura o alelo rs16891982*G (Tabelas 4 e 5), reforçando as associações encontradas para esse SNP.

Esta variante (rs16891982) também se localiza na 3'UTR da isoforma 3. Variações genéticas nessa região podem contribuir para a diversidade fenotípica na população humana através da regulação pós-transcricional por microRNAs (Bulik-Sullivan, Selitsky et al. 2013), pois estes são capazes de inibir a síntese proteica (parcial ou completamente) através da interação com a 3'UTR do RNAm (Kamaraj, Gopalakrishnan et al. 2014, Gulyaeva e Kushlinskiy 2016). Através da análise de

microRNAs, obtivemos os valores de ΔE (reflete a estabilidade da interação miRNA/mRNA), ΔS (reflete a especificidade da interação miRNA/mRNA) e ΔES (reflete um valor combinado de ambas as pontuações, ΔE e ΔS) para dois haplótipos que diferem quanto a esse SNP (alelos C e G). Quanto maior for a pontuação ΔE , maior a probabilidade do SNP influenciar a intensidade de interação entre o miRNA e o alvo; da mesma forma, quanto maior for a pontuação ΔS , maior a probabilidade do SNP influenciar a especificidade do miRNA em interagir somente com essa região e, por fim, quanto maior a pontuação ΔES , maior a probabilidade do SNP influenciar a intensidade e especificidade do miRNA em questão com a região alvo. De modo geral, dos 20 miRNA selecionados por apresentar maior ΔE , 12 apresentam ligação mais estável com o alelo G, enquanto que os outros 8 apresentam maior intensidade de ligação com o alelo C (Tabela 19). Cenário idêntico pode ser observado quanto ao ΔS e ΔES . Isso leva à conclusão que, apesar dos alelos do SNP rs16891982 apresentarem potencial de influenciar ligação com miRNAs quando tal SNP ocupar a 3'UTR do transcrito, não é possível estabelecer se transcritos carregando um dos alelos seriam efetivamente regulados de forma mais intensa por miRNAs do que transcritos portando o outro alelo.

Todas as variantes na 3'UTR da isoforma 2 que mostraram associação com alguma característica fenotípica (rs6894408, rs6874998, rs55746663, rs57959046, rs35407, rs7380538 e rs35406), estão em forte e completo desequilíbrio de ligação [$LOD \geq 2$ e $D' = 1$] com o rs16891982 (FIGURA 8). Dentre elas, os SNPs rs35407, rs35406, rs57959046 e rs55746663 já foram relatados em outros trabalhos em sítios de ligação de microRNAs (Bulik-Sullivan, Selitsky et al. 2013, Kamaraj, Gopalakrishnan et al. 2014). O alelo A do SNP rs35406 cria um sítio de ligação para 13 novos microRNAs, diminuindo a quantidade de produto proteico no citoplasma, prejudicando a produção de melanina (Kamaraj, Gopalakrishnan et al. 2014). Estes dados estão de acordo com a associação do alelo rs35406*A ($p = 0,0001$, OR = 15,73) com cabelos loiros que encontramos na nossa amostra (Tabela 13). Complementando este achado, o genótipo A/A foi associado no presente estudo a uma redução significativa do índice de melanina em comparação ao grupo com genótipo A/G, reforçando a associação encontrada do alelo rs35406*A a cabelos loiros (Tabela 17).

Na análise de microRNAs para os haplótipos dessa região, podemos observar que, em média, as interações mais estáveis e intensas envolvem os haplótipos utr4, utr5 e utr7, enquanto que os utr1, utr2, utr3 e utr6 são alvos secundários (Tabela 21). Enquanto que as associações envolvendo utr2 e utr3 com fenótipos escuros é coerente com estes achados de possível interação com miRNAs, as associações de utr1 com fenótipos claros e utr4, utr5 e utr7 com fenótipos escuros (Tabela 15) são, em

princípio, incompatíveis. Para uma avaliação definitiva é fundamental considerar a expressão ou não de cada miRNA nos melanócitos.

7.2.2 Gene TYR

7.2.2.1 REGIÃO PROMOTORA

Seis variantes (rs75741664, chr11:88907815, rs142881018, rs61759512, rs11824466 e rs55870815) da região promotora do gene *TYR* se mostraram associadas com fenótipos de pigmentação. Três delas (chr11:88907815, rs142881018 e rs55870815) foram consideradas monomórficas na população do presente estudo coletada aleatoriamente (Tabela 12), sendo que o alelo alternativo (G) da variante chr11:88907815 não havia sido descrito anteriormente e foi observado apenas na amostra brasileira enriquecida. O alelo alternativo da indel rs142881018 e do SNP rs55870815 está completamente ausente nas populações asiáticas e europeia, é raro em populações miscigenadas, como também observamos na população aqui representada, e tem baixa frequência em populações africanas (Tabela 24). Em concordância com essas informações, os alelos rs142881018*C e rs55870815*C se mostraram associados a pele escura ($p = 0,0001$, OR = 41,02) e cabelos pretos ($p = 0,0161$, OR = 20,92), respectivamente (Tabela 14).

Com relação as variantes polimórficas, podemos observar na Tabela 25 que os alelos alternativos rs75741664*T e rs61759512*A estão ausentes em todas as populações mundiais, exceto em africanos (rs75741664*T – 5,75% e rs61759512*A – 5,98%). Já o alelo rs11824466*T está ausente nas populações asiáticas (EAS e SAS), é raro nas populações EUR (0,20%) e AMR (0,72%) e possui maior frequência em AFR (9,83%). Estes dados concordam com suas associações encontradas para pele escura [rs75741664*T ($p = 0,0231$, OR = 10,63); rs61759512*A ($p = 0,024$, OR = 10,48) e rs11824466*T ($p = 0,0001$, OR = 25,08)] e com os resultados das análises de índice de melanina em que os genótipos contendo esses alelos (rs75741664*C/T, rs61759512*G/A e rs11824466*C/T) foram associados a um aumento significativo do índice de melanina em comparação aos genótipos homocigotos rs75741664*C/C, rs61759512*G/G e rs11824466*C/C (Tabela 18).

Tabela 24: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados na região promotora do gene *TYR*, nas populações mundiais e na população do presente estudo.

Polimorfismo	Alelo REF ALT	BRA (n=200)	EAS (n=504)	EUR (n=503)	AFR (n=661)	AMR (n=347)	SAS (n=489)
--------------	---------------------	----------------	----------------	----------------	----------------	----------------	----------------

rs75741664	C	0,9875	1,0000	1,0000	0,9425	1,0000	1,0000
	T	0,0125	-	-	0,0575	-	-
rs142881018	CCTGA	0,9900	1,0000	1,0000	0,9312	0,9942	1,0000
	C	0,0100	-	-	0,0688	0,0058	-
rs61759512	G	0,9825	1,0000	1,0000	0,9402	1,0000	1,0000
	A	0,0125	-	-	0,0598	-	-
rs11824466	C	0,9800	1,0000	0,9980	0,9017	0,9928	1,0000
	T	0,0200	-	0,0020	0,0983	0,0072	-
rs55870815	T	0,9950	1,0000	1,0000	0,9766	0,9971	1,0000
	C	0,0025	-	-	0,0234	0,0029	-

A busca por fatores de transcrição para a região promotora do gene *TYR* através da ferramenta HaploReg v4.1 (<http://archive.broadinstitute.org/mammals/haploreg/haploreg.php>) encontrou cinco fatores de transcrição com afinidade para o sítio onde se insere o SNP rs75741664 (BCL_disc4, HDAC2_disc6, Irf_known4, PRDM1_known1, Pax-5_disc3), quatro para o rs55870815 (FXR_1, Irf_disc3, Irf_Known4, ZEB1_Known4), dois para o rs11824466 (Pou2f2_known2, Pou2f2_known8) e um para a variante rs142881018 (PRDM1). Nenhum fator de transcrição se liga nos locais onde se encontram os SNPs rs61759512 e chr11:88907815. Com exceção do fator HDAC2_disc6 que tem uma afinidade ligeiramente maior (0,5) pelo alelo alternativo, todos os fatores encontrados para o SNP rs75741664 se ligam com maior afinidade ao alelo referência (rs75741664*C), sendo que o fator Irf_known4 tem a maior afinidade (11,9). Uma vez que este alelo se encontra associado a menor produção de melanina, as moléculas regulatórias que se ligam fortemente a ele provavelmente exercem função inibitória. Esta hipótese é corroborada pela informação de que os fatores BCL_disc4, Irf_known4 e PRDM1 foram encontrados expressos na pele e com função de repressão transcricional (<https://www.genecards.org/>). Dois fatores (Irf_disc3, Irf_Known4) relacionados ao SNP rs55870815 se ligam com maior afinidade na presença do alelo referência (rs55870815*T) e dois (FXR_1, ZEB1_Known4) na presença do alelo alternativo (rs55870815*C), o que consiste em achado de difícil interpretação. Do mesmo modo, considerando-se o SNP rs11824466, o fator de transcrição Pou2f2_known2 tem maior afinidade pelo alelo referência (rs11824466*C) do que pelo alelo alternativo (rs11824466*T), enquanto que o oposto ocorre para o fator Pou2f2_known8. E finalmente, o fator encontrado para a variante rs142881018 (PRDM1), se liga com afinidade 1,9 vezes maior ao alelo alternativo (rs142881018*C) do que ao alelo referência (rs142881018*CCTGA).

Alguns haplótipos da região promotora do gene *TYR* estão associados a fenótipos de pigmentação escuros: os haplótipos “promo7” ($p = 0,0014$, OR = 14,06), “promo8” ($p = 9,95 \times 10^{-06}$, OR = 23,27), “promo9” ($p = 0,0173$, OR = 11,87) e “promo10” ($p =$

0,0014, OR = 14,06) se mostraram associados a pele escura e o haplótipo “promo15” ($p = 0,0172$, OR = 20,35) a cabelos pretos (Tabela 16). Todos os haplótipos mencionados, possuem ao menos o alelo chr11:88907815*A (vale ressaltar que o alelo chr11:88907815*G foi associado com cabelos loiros). Além disso, com exceção da promo7, os demais haplótipos apresentam um ou dois outros alelos mencionados acima (rs75741664, rs142881018, rs61759512, rs11824466 e rs55870815) individualmente associado a fenótipos escuros.

7.2.2.2 REGIÃO CODIFICANTE

Foram encontradas oito variantes associadas com características de pigmentação na região codificante do gene *TYR*, seis localizadas em íntrons (rs11018530, rs67512950, rs12804012, rs75375648, rs113438665 e rs1393350) e duas em éxons (rs1042602 e rs1126809). Também foi encontrado um SNP (rs28521275) localizado na região *downstream* desse gene cujo genótipo rs28521275*C/C se mostrou associado com cabelos ruivos ($p = 0,0014$, OR = 10,80) (Tabela 14).

De todos os sítios mencionados acima, apenas o SNP rs75375648, localizado no íntron dois, foi considerado monomórfico na população estudada (Tabela 12). Seu alelo alternativo (rs75375648*C) também é raro em populações miscigenadas (AMR - 0,86%) e completamente ausente em europeus e asiáticos, porém com maior frequência em africanos (11,57%) (Tabela 25). Este alelo se mostrou associado a pele escura ($p = 0,0017$, OR = 29,32) no presente trabalho (Tabela 14), o que é condizente com os resultados das análises de índice de melanina em que o genótipo contendo esse alelo rs75375648*T/C foi associado a um aumento significativo do índice de melanina em comparação ao genótipo homocigoto referência rs75375648*T/T (Tabela 18).

Dentre as variantes polimórficas localizadas na região intrônica, o genótipo rs1393350*A/A mostrou associações com olhos azuis ($p = 0,0253$, OR = 13,06) e cabelos castanho-claros ($p = 0,0019$, OR = 16,07), mas seus alelos não apresentaram associações significativas com nenhuma das características avaliadas (Tabela 14). Estas associações estão de acordo com os resultados das análises de índice de melanina na pele, em que o genótipo rs1393350*G/G foi associado a um aumento significativo do índice de melanina em comparação aos grupos com os genótipos rs1393350*A/A e rs1393350*G/A (Tabela 18). Também podemos observar concordância com a maior frequência do alelo rs1393350*A na população europeia (24,35%), em que os indivíduos normalmente apresentam fenótipos mais claros, e menor frequência na população africana (0,83%) (Tabela 25). É interessante observar que as populações miscigenadas (BRA – 18% e AMR – 12,25%) apresentam

frequências intermediárias em relação a europeia e africana, com maior tendência para europeia, realçando as características de miscigenação e colonização. Este SNP faz parte do conjunto de SNPs do sistema HirisPlex para predição das cores de olhos e cabelos (Walsh, Liu et al. 2013) e já foi associado com habilidade de bronzeamento em euro-americanos em um GWAS (*Genome-Wide Association Study*) (Nan, Kraft et al. 2009). Neste mesmo trabalho, este SNP se mostrou em forte desequilíbrio de ligação com o SNP exônico rs1126809 ($D' = 1$ e $r^2 = 0,86$). Ambos SNPs também estão em forte desequilíbrio de ligação ($LOD \geq 2$ e $D' = 0,98$) no presente estudo (FIGURA 9).

Tabela 25: Frequências alélicas dos SNPs associados a algum dos fenótipos de pigmentação avaliados, localizados na região codificante e *downstream* do gene *TYR*, nas populações mundiais e na população do presente estudo.

Polimorfismo	Alelo REF ALT	BRA (n=200)	EAS (n=504)	EUR (n=503)	AFR (n=661)	AMR (n=347)	SAS (n=489)
rs1042602	C	0,6750	0,9990	0,6282	0,9879	0,7622	0,9366
	A	0,3250	0,0010	0,3718	0,0121	0,2378	0,0634
rs11018530	C	0,8450	0,8750	0,8797	0,5431	0,8487	0,8344
	G	0,1525	0,1250	0,1203	0,4569	0,1513	0,1656
rs67512950	ACTT	0,8600	0,8750	0,8817	0,6165	0,8516	0,8344
	A	0,1400	0,1250	0,1183	0,3835	0,1484	0,1656
rs12804012	G	0,8325	0,8740	0,8817	0,4614	0,8401	0,7853
	A	0,1650	0,1260	0,1183	0,5386	0,1599	0,2147
rs75375648	T	0,9900	1,0000	1,0000	0,8843	0,9914	1,0000
	C	0,0075	-	-	0,1157	0,0086	-
rs113438665	T	0,9850	1,0000	0,9990	0,9236	0,9899	1,0000
	C	0,0125	-	0,0010	0,0764	0,0101	-
rs1393350	G	0,8200	0,9970	0,7565	0,9917	0,8775	0,9458
	A	0,1800	0,0030	0,2435	0,0083	0,1225	0,0542
rs1126809	G	0,8025	0,9990	0,7475	0,9909	0,8746	0,9458
	A	0,1850	0,0010	0,2525	0,0091	0,1254	0,0542
rs28521275	C	0,4475	0,5139	0,5229	0,1036	0,4625	0,5348
	T	0,4775	0,4861	0,4771	0,8964	0,5375	0,4652

O SNP rs1126809 está localizado no éxon 4 e determina uma mutação não-sinônima, causando a substituição do aminoácido arginina para glutamina. Da mesma forma que para o SNP rs1393350, foram encontradas apenas associações significativas entre seu genótipo rs1126809*A/A e as cores de olhos azuis ($p = 0,0249$, OR = 13,18) e cabelos castanho-claros ($p = 0,0056$, OR = 13,44) (Tabela 14). Esses dados são congruentes com a maior frequência do alelo rs1126809*A na população europeia (25,25%) (Tabela 25) e também com trabalho realizado no Brasil em que o alelo rs1126809*G se mostrou associado com características de pigmentação de pele mais escuras (Cerqueira, Hunemeier et al. 2014). Quando observamos os resultados

das análises de índice de melanina na pele, os genótipos rs1126809*A/A e rs1126809*G/A mostram uma diminuição significativa do índice de melanina em comparação ao genótipo homocigoto rs1126809*G/G (Tabela 18).

Outro SNP localizado em éxon que se mostrou associado com os fenótipos avaliados foi o rs1042602. Sua mutação também é não-sinônima causando a substituição do aminoácido serina por tirosina. Na presente amostra, apenas o genótipo rs1042602*A/A apresentou associação significativa com a cor de pele clara ($p = 0,0033$, OR = 11,00) (Tabela 14). Os resultados das análises de índice de melanina refletem a associação encontrada já que o genótipo rs1042602*C/C mostra um aumento significativo do índice de melanina em comparação aos genótipos rs1042602*C/A e rs1042602*A/A (Tabela 18). Em relação à frequência dessa variante, podemos notar que o alelo alternativo (rs1042602*A) é bem mais frequente na população europeia (37,18%) do que nas populações africana (1,21%) e asiáticas (EAS – 0,1% e SAS – 6,34%), porém é melhor distribuído nas populações miscigenadas (BRA – 32,5% e AMR – 23,78%) (Tabela 25). Em trabalho GWAS sua frequência foi observada em concordância com as reportadas acima em europeus (37%), porém sem apresentar associação significativa com a pigmentação da pele (Candille, Absher et al. 2012).

Ao analisarmos as associações encontradas para os haplótipos inferidos para essa região do gene *TYR*, podemos notar que os haplótipos “cds5” ($p = 1,71 \times 10^{-05}$, OR = 21,2549), “cds7” ($p = 0,0061$, OR = 23,7042) e “cds9” ($p = 0,0017$, OR = 29,2464) estão associados a pele escura e o haplótipo “cds12” ($p = 0,015$, OR = 21,61) a ausência de sardas (Tabela 16). Quando nos atentamos a composição desses haplótipos, observa-se ao menos a presença dos alelos rs1042602, rs1393350 e rs1126809 previamente associados a fenótipos escuros (Tabela 10), reforçando as associações encontradas para eles.

Outras três variantes intrônicas mostraram as associações mais fortes encontradas com as características avaliadas. Os genótipos rs11018530*C/C ($p = 4,96 \times 10^{-06}$, OR = 13,15), rs67512950*ACTT/ACTT ($p = 2,58 \times 10^{-05}$, OR = 11,28) e rs12804012*G/G ($p = 3,72 \times 10^{-06}$, OR = 10,28) estão associados a olhos azuis e os genótipos rs11018530*G/G ($p = 3,89 \times 10^{-05}$, OR = 15,55) e rs12804012*A/A ($p = 1,40 \times 10^{-06}$, OR = 20,29) a pele escura (Tabela 14). Os SNPs rs11018530 e rs12804012 estão em forte e completo desequilíbrio de ligação [$LOD \geq 2$ e $D' = 1$] (FIGURA 9), o que pode indicar que eventualmente apenas um deles apresente relevância funcional. Os resultados das análises de índice de melanina corroboram os dados das associações já que os genótipos que possuem os alelos referência dessas variantes (rs11018530*C/C e rs11018530*C/G; rs67512950*ACTT/ACTT e rs67512950*ACTT/A;

rs12804012*G/G e rs12804012*G/A) mostram uma diminuição significativa do índice de melanina em comparação aos grupos com os genótipos homocigoto alternativo rs11018530*G/G, rs67512950*A/A e rs12804012*A/A (Tabela 18). Outros dados que fortalecem esses resultados são as frequências dos alelos desses SNPs nas populações mundiais (Tabela 25). Os alelos referência (rs11018530*C, rs67512950*ACTT e rs12804012*G), cujos genótipos homocigotos se mostraram associados com a cor de olhos azuis, são mais frequentes na população europeia (87,97%, 88,17% e 88,17%; respectivamente) do que na população africana (54,31%, 61,65% e 46,14%; respectivamente).

7.3 Considerações finais

Foram encontradas evidências de que diferentes pressões seletivas estejam atuando em genes de pigmentação em diferentes populações (Sturm 2009). Algumas das possíveis pressões seletivas agindo sobre a pigmentação da pele em ambientes com alto ou baixo índice de radiação UV incluem a necessidade de síntese de vitamina D, a proteção contra a fotólise do folato, manutenção da pele como uma barreira contra perda de água e, ao mesmo tempo, a primeira linha de defesa do sistema imune inato, assim como uma proteção contra o câncer de pele e outras doenças (Sturm and Duffy 2012).

Muitas evidências apontam para uma seleção da pigmentação escura da pele como uma medida de proteção contra os efeitos deletérios (por ex., fotólise do folato) da radiação UV em localidades tropicais. Em contrapartida, como a vitamina D3 só é sintetizada quando a radiação UV penetra na pele, a pigmentação clara é uma proteção contra as graves consequências no desenvolvimento (anormalidades no crescimento, raquitismo e osteoporose) que a falta de tal vitamina pode ter (Yuen and Jablonski 2010, Sturm and Duffy 2012). Dessa forma, podemos observar que uma dupla pressão seletiva, fotoproteção e níveis de vitamina D, criou a necessidade de padrões regionais de peles mais e menos pigmentadas, o que desenvolveu a grande variedade fenotípica encontrada na pigmentação nos humanos modernos.

Por se tratar de uma população altamente miscigenada, os polimorfismos de ambos os genes apresentam distribuições de frequências alélicas muito parecidas com a de populações miscigenadas que ocupam o continente americano (AMR) ou intermediária em relação as populações europeias (EUR) e africanas (AFR). Também foram encontrados alelos de polimorfismos associados com características de pigmentação clara ou escura que apresentaram maiores frequências nas populações EUR e AFR, respectivamente. Em um estudo de associação convencional, tais

associações poderiam ser tratadas como espúrias, isto é, relacionadas a ancestralidade, não refletindo necessariamente a diversidade das vias que determinam fenótipos pigmentares. Porém, podemos observar que a maioria desses alelos está localizada em sítios de ligação de fatores de transcrição ou miRNAs que funcionam como repressores ou ativadores transcricionais, o que atesta a significância biológica de tais associações.

Determinar a diversidade e associações fenotípicas dos alelos e haplótipos em genes pigmentares de sabida importância numa população miscigenada como a brasileira é de grande interesse na expansão do conhecimento acerca de como essas variações na pigmentação e as forças evolutivas podem interagir. A avaliação de interações epistáticas entre variantes dos genes *SLC45A2* e *TYR* no contexto de características pigmentares poderia conduzir à compreensão mais detalhada do papel de ambos na pigmentação em humanos, acarretando na identificação de possíveis interações que venham a refletir o papel desempenhado pela proteína MATP no transporte de tirosinase, com consequente impacto em variações fenotípicas da pele, olhos e cabelos.

8. CONCLUSÕES

A variação na distribuição geográfica das frequências alélicas da maior parte das variantes desses genes e as frequências observadas na presente amostra são consistentes com a existência de mistura inter-étnica na população de Ribeirão Preto.

Foram encontradas 25 variantes (das 58 analisadas) polimórficas para o gene *SLC45A2* e 19 (das 42 analisadas) para o gene *TYR*. Nenhum sítio de variação foi encontrado na região 5'UTR do gene *SLC45A2* e nas regiões 5'UTR e 3'UTR do gene *TYR*. As regiões mais diversas em ambos os genes foram as regiões intrônicas.

Vinte e oito variantes do gene *SLC45A2* e 15 do gene *TYR*, mostraram associações significativas com as características de pigmentação analisadas. Oitenta e uma associações com fenótipos pigmentares foram encontradas para os haplótipos distribuídos nas regiões promotora, codificante e 3'UTR do gene *SLC45A2*. Quinze associações com fenótipos pigmentares foram encontradas para os haplótipos distribuídos nas regiões promotora, codificante e *downstream* do gene *TYR*.

O SNP rs16891982 (Leu374Phe) mostrou as mais fortes associações encontradas no presente estudo, corroborando diversos relatos na literatura. Em relação a associações de haplótipos codificantes das isoformas 1 e 2, a presença de um ou do outro alelo foi determinante nas associações haplotípicas com fenótipos claros ou escuros. Em relação a sua presença na 3'UTR da isoforma 3, apesar de ambos os

alelos apresentarem potencial de influenciar ligação com miRNAs, não é possível estabelecer se transcritos carregando um dos alelos seriam efetivamente regulados de forma mais intensa por miRNAs do que transcritos portando o outro alelo.

Foram encontradas evidências de interação de fatores de transcrição com sítios polimórficos nas regiões promotora de ambos os genes e de microRNAs na região 3'UTR das três isoformas do gene *SLC45A2*. Estas interações, envolvendo funções de ativação ou repressão da expressão gênica são, em sua maioria, consistentes com as associações relatadas.

Foram encontrados blocos em desequilíbrio de ligação envolvendo polimorfismos em todas as regiões em ambos os genes. Os SNPs rs28777, rs183671 e rs26722 do gene *SLC45A2* pertencem ao mesmo bloco e estão em forte desequilíbrio com o SNP rs16891982, porém as fortes associações encontradas para eles indicam que seus alelos podem influenciar a magnitude das associações. Já os SNPs rs11018530 e rs12804012 do gene *TYR* estão em forte e completo desequilíbrio de ligação, sugerindo que apenas um deles apresente relevância funcional.

De modo geral, os resultados aqui encontrados reafirmam a importância desempenhada pelos genes *SLC45A2* e *TYR* na geração da diversidade de fenótipos de pigmentação.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- Ainger, S. A., K. Jagirdar, K. J. Lee, H. P. Soyer and R. A. Sturm (2017). "Skin Pigmentation Genetics for the Clinic." *Dermatology* **233**(1): 1-15.
- Albert, F. W. and L. Kruglyak (2015). "The role of regulatory variation in complex traits and disease." *Nat Rev Genet* **16**(4): 197-212.
- Andrews, S. (2010). "FastQC: a quality control tool for high throughput sequence data." from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bae-Harboe, Y. S. and H. Y. Park (2012). "Tyrosinase: a central regulatory protein for cutaneous pigmentation." *J Invest Dermatol* **132**(12): 2678-2680.
- Banovich, N. E., X. Lan, G. McVicker, B. van de Geijn, J. F. Degner, J. D. Blischak, J. Roux, J. K. Pritchard and Y. Gilad (2014). "Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels." *PLoS Genet* **10**(9): e1004663.
- Barrett, J. C., B. Fry, J. Maller and M. J. Daly (2005). "Haploview: analysis and visualization of LD and haplotype maps." *Bioinformatics* **21**(2): 263-265.
- Beleza, S., N. A. Johnson, S. I. Candille, D. M. Absher, M. A. Coram, J. Lopes, J. Campos, Araujo, II, T. M. Anderson, B. J. Vilhjalmsson, M. Nordborg, E. S. A. Correia, M. D. Shriver, J. Rocha, G. S. Barsh and H. Tang (2013). "Genetic architecture of skin and eye color in an African-European admixed population." *PLoS Genet* **9**(3): e1003372.
- Beleza, S., A. M. Santos, B. McEvoy, I. Alves, C. Martinho, E. Cameron, M. D. Shriver, E. J. Parra and J. Rocha (2013). "The timing of pigmentation lightening in Europeans." *Mol Biol Evol* **30**(1): 24-35.

Bewick, V., L. Cheek and J. Ball (2004). "Statistics review 10: further nonparametric methods." *Crit Care* **8**(3): 196-199.

Bikle, D. D. (2014). "Vitamin D metabolism, mechanism of action, and clinical applications." *Chem Biol* **21**(3): 319-329.

Bin, B. H., J. Bhin, S. H. Yang, M. Shin, Y. J. Nam, D. H. Choi, D. W. Shin, A. Y. Lee, D. Hwang, E. G. Cho and T. R. Lee (2015). "Membrane-Associated Transporter Protein (MATP) Regulates Melanosomal pH and Influences Tyrosinase Activity." *PLoS One* **10**(6): e0129273.

Branicki, W., U. Brudnik, J. Draus-Barini, T. Kupiec and A. Wojas-Pelc (2008). "Association of the SLC45A2 gene with physiological human hair colour variation." *J Hum Genet* **53**(11-12): 966-971.

Branicki, W., F. Liu, K. van Duijn, J. Draus-Barini, E. Pospiech, S. Walsh, T. Kupiec, A. Wojas-Pelc and M. Kayser (2011). "Model-based prediction of human hair color using DNA variants." *Hum Genet* **129**(4): 443-454.

Bulik-Sullivan, B., S. Selitsky and P. Sethupathy (2013). "Prioritization of genetic variants in the microRNA regulome as functional candidates in genome-wide association studies." *Hum Mutat* **34**(8): 1049-1056.

Busch, A., A. S. Richter and R. Backofen (2008). "IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions." *Bioinformatics* **24**(24): 2849-2856.

Calvo, P. A., D. W. Frank, B. M. Bieler, J. F. Berson and M. S. Marks (1999). "A cytoplasmic sequence in human tyrosinase defines a second class of di-leucine-based sorting signals for late endosomal and lysosomal delivery." *J Biol Chem* **274**(18): 12780-12789.

Candille, S. I., D. M. Absher, S. Beleza, M. Bauchet, B. McEvoy, N. A. Garrison, J. Z. Li, R. M. Myers, G. S. Barsh, H. Tang and M. D. Shriver (2012). "Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four European populations." *PLoS One* **7**(10): e48294.

Cannell, J. J., B. W. Hollis, M. Zasloff and R. P. Heaney (2008). "Diagnosis and treatment of vitamin D deficiency." *Expert Opin Pharmacother* **9**(1): 107-118.

Cartegni, L., S. L. Chew and A. R. Krainer (2002). "Listening to silence and understanding nonsense: exonic mutations that affect splicing." *Nat Rev Genet* **3**(4): 285-298.

Castelli, E. C., C. T. Mendes-Junior, A. Sabbagh, I. O. Porto, A. Garcia, J. Ramalho, T. H. Lima, J. D. Massaro, F. C. Dias, C. V. Collares, V. Jamonneau, B. Bucheton, M. Camara and E. A. Donadi (2015). "HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples." *Hum Immunol* **76**(12): 945-953.

Catalanotto, C., C. Cogoni and G. Zardo (2016). "MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions." *Int J Mol Sci* **17**(10).

Cerqueira, C. C., T. Hunemeier, J. Gomez-Valdes, V. Ramallo, C. D. Volasko-Krause, A. A. Barbosa, P. Vargas-Pinilla, R. C. Dornelles, D. Longo, F. Rothhammer, G. Bedoya, S. Canizales-Quinteros, V. Acuna-Alonzo, C. Gallo, G. Poletti, R. Gonzalez-Jose, F. M. Salzano, S. M. Callegari-Jacques, L. Schuler-Faccini, A. Ruiz-Linares, M. Catira Bortolini and C. for (2014). "Implications of the admixture process in skin color molecular assessment." *PLoS One* **9**(5): e96886.

Chamary, J. V., J. L. Parmley and L. D. Hurst (2006). "Hearing silence: non-neutral evolution at synonymous sites in mammals." *Nat Rev Genet* **7**(2): 98-108.

Chen, N., J. R. Wang, L. Huang, Y. Yang, Y. M. Jiang, X. J. Guo, Y. Z. He and Y. H. Zhou (2015). "Significant association of alpha-methylacyl-CoA racemase gene

polymorphisms with susceptibility to prostate cancer: a meta-analysis." Asian Pac J Cancer Prev **16**(5): 1857-1863.

Cichorek, M., M. Wachulska, A. Stasiewicz and A. Tyminska (2013). "Skin melanocytes: biology and development." Postepy Dermatol Alergol **30**(1): 30-41.

Clark, P. U., A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark, B. Wohlfarth, J. X. Mitrovica, S. W. Hostetler and A. M. McCabe (2009). "The Last Glacial Maximum." Science **325**(5941): 710-714.

Comeron, J. M. (2004). "Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence." Genetics **167**(3): 1293-1304.

Costin, G. E. and V. J. Hearing (2007). "Human skin pigmentation: melanocytes modulate skin color in response to stress." FASEB J **21**(4): 976-994.

Costin, G. E., J. C. Valencia, W. D. Vieira, M. L. Lamoreux and V. J. Hearing (2003). "Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4." J Cell Sci **116**(Pt 15): 3203-3212.

Cullinane, A. R., T. Vilboux, K. O'Brien, J. A. Curry, D. M. Maynard, H. Carlson-Donohoe, C. Ciccone, N. C. S. Program, T. C. Markello, M. Gunay-Aygun, M. Huizing and W. A. Gahl (2011). "Homozygosity mapping and whole-exome sequencing to detect SLC45A2 and G6PC3 mutations in a single patient with oculocutaneous albinism and neutropenia." J Invest Dermatol **131**(10): 2017-2025.

D'Mello, S. A., G. J. Finlay, B. C. Baguley and M. E. Askarian-Amiri (2016). "Signaling Pathways in Melanogenesis." Int J Mol Sci **17**(7).

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernysky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler and M. J. Daly (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nat Genet **43**(5): 491-498.

Dooley, C. M., H. Schwarz, K. P. Mueller, A. Mongera, M. Konantz, S. C. Neuhaus, C. Nusslein-Volhard and R. Geisler (2013). "Slc45a2 and V-ATPase are regulators of melanosomal pH homeostasis in zebrafish, providing a mechanism for human pigment evolution and disease." Pigment Cell Melanoma Res **26**(2): 205-217.

Duval, C., C. Cohen, C. Chagnoleau, V. Flouret, E. Bourreau and F. Bernerd (2014). "Key regulatory role of dermal fibroblasts in pigmentation as demonstrated using a reconstructed skin model: impact of photo-aging." PLoS One **9**(12): e114182.

Excoffier, L. and H. E. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Mol Ecol Resour **10**(3): 564-567.

Fernandez, L. P., R. L. Milne, G. Pita, J. A. Aviles, P. Lazaro, J. Benitez and G. Ribas (2008). "SLC45A2: a novel malignant melanoma-associated gene." Hum Mutat **29**(9): 1161-1167.

Ferragina, P., G. Manzini, V. Mäkinen and G. Navarro (2004). An Alphabet-Friendly FM-Index. String Processing and Information Retrieval, Springer, Berlin, Heidelberg: 150-160.

Ferreira, L. B., C. T. Mendes-Junior, C. E. Wiezel, M. R. Luizon and A. L. Simoes (2006). "Genomic ancestry of a sample population from the state of Sao Paulo, Brazil." Am J Hum Biol **18**(5): 702-705.

Fesenko, D. O., A. V. Chudinov, S. A. Surzhikov and A. S. Zasedatelev (2016). "Biochip-Based Genotyping Assay for Detection of Polymorphisms in Pigmentation

Genes Associated with Cutaneous Melanoma." *Genet Test Mol Biomarkers* **20**(4): 208-212.

Fitzpatrick, T. B. (1988). "The validity and practicality of sun-reactive skin types I through VI." *Arch Dermatol* **124**(6): 869-871.

Fracasso, N. C. A., E. S. de Andrade, C. E. V. Wiesel, C. C. F. Andrade, L. R. Zanao, M. S. da Silva, L. A. Marano, E. A. Donadi, C. C. E, A. L. Simoes and C. T. Mendes-Junior (2017). "Haplotypes from the SLC45A2 gene are associated with the presence of freckles and eye, hair and skin pigmentation in Brazil." *Leg Med (Tokyo)* **25**: 43-51.

Friedman, R. C., K. K. Farh, C. B. Burge and D. P. Bartel (2009). "Most mammalian mRNAs are conserved targets of microRNAs." *Genome Res* **19**(1): 92-105.

Geer, L. Y., A. Marchler-Bauer, R. C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi and S. H. Bryant (2010). "The NCBI BioSystems database." *Nucleic Acids Res* **38**(Database issue): D492-496.

Gegotek, A. and E. Skrzydlewska (2015). "The role of transcription factor Nrf2 in skin cells metabolism." *Arch Dermatol Res* **307**(5): 385-396.

Genomes Project, C., A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korb, J. L. Marchini, S. McCarthy, G. A. McVean and G. R. Abecasis (2015). "A global reference for human genetic variation." *Nature* **526**(7571): 68-74.

Graf, J., R. Hodgson and A. van Daal (2005). "Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation." *Hum Mutat* **25**(3): 278-284.

Gulyaeva, L. F. and N. E. Kushlinskiy (2016). "Regulatory mechanisms of microRNA expression." *J Transl Med* **14**(1): 143.

Han, J., P. Kraft, H. Nan, Q. Guo, C. Chen, A. Qureshi, S. E. Hankinson, F. B. Hu, D. L. Duffy, Z. Z. Zhao, N. G. Martin, G. W. Montgomery, N. K. Hayward, G. Thomas, R. N. Hoover, S. Chanock and D. J. Hunter (2008). "A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation." *PLoS Genet* **4**(5): e1000074.

Hart, A. (2001). "Mann-Whitney test is not just a test of medians: differences in spread can be important." *BMJ* **323**(7309): 391-393.

Hearing, V. J. (2011). "Determination of melanin synthetic pathways." *J Invest Dermatol* **131**(E1): E8-E11.

<http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/> (2018). 1000 Genomes Browser.

Hudjashov, G., R. Vellems and T. Kivisild (2013). "Global patterns of diversity and selection in human tyrosinase gene." *PLoS One* **8**(9): e74307.

Hull, J., S. Campino, K. Rowlands, M. S. Chan, R. R. Copley, M. S. Taylor, K. Rockett, G. Elvidge, B. Keating, J. Knight and D. Kwiatkowski (2007). "Identification of common genetic variation that modulates alternative splicing." *PLoS Genet* **3**(6): e99.

Ibarrola-Villava, M., L. P. Fernandez, S. Alonso, M. D. Boyano, M. Pena-Chilet, G. Pita, J. A. Aviles, M. Mayor, C. Gomez-Fernandez, B. Casado, M. Martin-Gonzalez, N. Izagirre, C. De la Rua, A. Asumendi, G. Perez-Yarza, Y. Arroyo-Berdugo, E. Boldo, R. Lozoya, A. Torrijos-Aguilar, A. Pitarch, G. Pitarch, J. M. Sanchez-Motilla, F. Valcuende-Cavero, G. Tomas-Cabedo, G. Perez-Pastor, J. L. Diaz-Perez, J. Gardeazabal, I. Martinez de Lizarduy, A. Sanchez-Diez, C. Valdes, A. Pizarro, M. Casado, G. Carretero, R. Botella-Estrada, E. Nagore, P. Lazaro, A. Lluch, J. Benitez, C. Martinez-Cadenas and G. Ribas (2011). "A customized pigmentation SNP array identifies a novel SNP associated with melanoma predisposition in the SLC45A2 gene." *PLoS One* **6**(4): e19271.

Ibarrola-Villava, M., H. H. Hu, M. Guedj, L. P. Fernandez, V. Descamps, N. Basset-Seguín, M. Bagot, A. Bensussan, P. Saiag, M. C. Fagnoli, K. Peris, J. A. Aviles, A. Lluch, G. Ribas and N. Soufir (2012). "MC1R, SLC45A2 and TYR genetic variants involved in melanoma susceptibility in southern European populations: results from a meta-analysis." *Eur J Cancer* **48**(14): 2183-2191.

IBGE, I. B. d. G. e. E.-. (2007). *Brasil: 500 anos de povoamento*. Rio de Janeiro, IBGE.

IBGE, I. B. d. G. e. E.-. (2010). 2018, from <http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm>.

Jablonski, N. G. (2004). "The evolution of human skin and skin color." *Annual Review of Anthropology* **33**: 585-623.

Jablonski, N. G. (2012). *Living Color - The Biological and Social Meaning of Skin Color*. Berkeley and Los Angeles, California, University of California Press.

Jablonski, N. G. and G. Chaplin (2000). "The evolution of human skin coloration." *J Hum Evol* **39**(1): 57-106.

Jablonski, N. G. and G. Chaplin (2010). "Human skin pigmentation as an adaptation to UV radiation." *Proceedings of the National Academy of Sciences of the United States of America* **107**: 8962-8968.

Jacobs, L. C., F. Liu, L. M. Pardo, A. Hofman, A. G. Uitterlinden, M. Kayser and T. Nijsten (2015). "IRF4, MC1R and TYR genes are risk factors for actinic keratosis independent of skin color." *Hum Mol Genet* **24**(11): 3296-3303.

Jagirdar, K., D. J. Smit, S. A. Ainger, K. J. Lee, D. L. Brown, B. Chapman, Z. Zhen Zhao, G. W. Montgomery, N. G. Martin, J. L. Stow, D. L. Duffy and R. A. Sturm (2014). "Molecular analysis of common polymorphisms within the human Tyrosinase locus and genetic association with pigmentation traits." *Pigment Cell Melanoma Res* **27**(4): 552-564.

Jani, R. A., L. K. Purushothaman, S. Rani, P. Bergam and S. R. Setty (2015). "STX13 regulates cargo delivery from recycling endosomes during melanosome biogenesis." *J Cell Sci* **128**(17): 3263-3276.

John, B., A. J. Enright, A. Aravin, T. Tuschl, C. Sander and D. S. Marks (2004). "Human MicroRNA targets." *PLoS Biol* **2**(11): e363.

Jonnalagadda, M., H. Norton, S. Ozarkar, S. Kulkarni and R. Ashma (2016). "Association of genetic variants with skin pigmentation phenotype among populations of west Maharashtra, India." *Am J Hum Biol* **28**(5): 610-618.

Kamaraj, B., C. Gopalakrishnan and R. Purohit (2014). "In silico analysis of miRNA-mediated gene regulation in OCA and OA genes." *Cell Biochem Biophys* **70**(3): 1923-1932.

Kamaraj, B. and R. Purohit (2014). "Mutational analysis of oculocutaneous albinism: a compact review." *Biomed Res Int* **2014**: 905472.

Kawakami, A. and D. E. Fisher (2011). "Key discoveries in melanocyte development." *J Invest Dermatol* **131**(E1): E2-4.

Kayser, M. (2015). "Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes." *Forensic Sci Int Genet* **18**: 33-48.

Kehdy, F. S., M. H. Gouveia, M. Machado, W. C. Magalhaes, A. R. Horimoto, B. L. Horta, R. G. Moreira, T. P. Leal, M. O. Scliar, G. B. Soares-Souza, F. Rodrigues-Souza, G. S. Araujo, R. Zamudio, H. P. Sant Anna, H. C. Santos, N. E. Duarte, R. L. Fiaccone, C. A. Figueiredo, T. M. Silva, G. N. Costa, S. Beleza, D. E. Berg, L. Cabrera, G. Debortoli, D. Duarte, S. Ghirotto, R. H. Gilman, V. F. Goncalves, A. R. Marrero, Y. C. Muniz, H. Weissensteiner, M. Yeager, L. C. Rodrigues, M. L. Barreto, M. F. Lima-Costa, A. C. Pereira, M. R. Rodrigues, E. Tarazona-Santos and E. P. C. Brazilian

(2015). "Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations." *Proc Natl Acad Sci U S A* **112**(28): 8696-8701.

Kondo, T., T. Namiki, S. G. Coelho, J. C. Valencia and V. J. Hearing (2015). "Oculocutaneous albinism: developing novel antibodies targeting the proteins associated with OCA2 and OCA4." *J Dermatol Sci* **77**(1): 21-27.

Korner, A. and J. Pawelek (1982). "Mammalian tyrosinase catalyzes three reactions in the biosynthesis of melanin." *Science* **217**(4565): 1163-1165.

Lai, X., H. J. Wichers, M. Soler-Lopez and B. W. Dijkstra (2018). "Structure and Function of Human Tyrosinase and Tyrosinase-Related Proteins." *Chemistry* **24**(1): 47-55.

Lassmann, T., Y. Hayashizaki and C. O. Daub (2011). "SAMStat: monitoring biases in next generation sequencing data." *Bioinformatics* **27**(1): 130-131.

Leite, T. K., R. M. Fonseca, N. M. de Franca, E. J. Parra and R. W. Pereira (2011). "Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings." *PLoS One* **6**(11): e27162.

Li, C. F., F. M. Fang, J. Lan, J. W. Wang, H. J. Kung, L. T. Chen, T. J. Chen, S. H. Li, Y. H. Wang, H. C. Tai, S. C. Yu and H. Y. Huang (2014). "AMACR amplification in myxofibrosarcomas: a mechanism of overexpression that promotes cell proliferation with therapeutic relevance." *Clin Cancer Res* **20**(23): 6141-6152.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." *Bioinformatics* **25**(14): 1754-1760.

Lima, T. H. A., R. V. Buttura, E. A. Donadi, L. C. Veiga-Castelli, C. T. Mendes-Junior and E. C. Castelli (2016). "HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample." *Hum Immunol* **77**(10): 841-853.

Liu, F., M. Visser, D. L. Duffy, P. G. Hysi, L. C. Jacobs, O. Lao, K. Zhong, S. Walsh, L. Chaitanya, A. Wollstein, G. Zhu, G. W. Montgomery, A. K. Henders, M. Mangino, D. Glass, V. Bataille, R. A. Sturm, F. Rivadeneira, A. Hofman, I. W. F. van, A. G. Uitterlinden, R. J. Palstra, T. D. Spector, N. G. Martin, T. E. Nijsten and M. Kayser (2015). "Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up." *Hum Genet* **134**(8): 823-835.

Liu, F., B. Wen and M. Kayser (2013). "Colorful DNA polymorphisms in humans." *Semin Cell Dev Biol* **24**(6-7): 562-575.

Lopez, S., O. Garcia, I. Yurrebaso, C. Flores, M. Acosta-Herrera, H. Chen, J. Gardeazabal, J. M. Careaga, M. D. Boyano, A. Sanchez, J. A. Raton-Nieto, A. Sevilla, I. Smith-Zubiaga, A. G. de Galdeano, C. Martinez-Cadenas, N. Izagirre, C. de la Rua and S. Alonso (2014). "The interplay between natural selection and susceptibility to melanoma on allele 374F of SLC45A2 gene in a South European population." *PLoS One* **9**(8): e104367.

Marano, L. A. (2011). Análise de polimorfismos do gene MC1R associados a fenótipos humanos de pigmentação na população Brasileira, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP).

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." *EMBnet.journal* **17**(1): 10.

Matoulkova, E., E. Michalova, B. Vojtesek and R. Hrstka (2012). "The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells." *RNA Biol* **9**(5): 563-576.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome

Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

Mengel-From, J., T. H. Wong, N. Morling, J. L. Rees and I. J. Jackson (2009). "Genetic determinants of hair and eye colours in the Scottish and Danish populations." BMC Genet **10**: 88.

Miller, S. A., D. D. Dykes and H. F. Polesky (1988). "A simple salting out procedure for extracting DNA from human nucleated cells." Nucleic Acids Res **16**(3): 1215.

Mishra, S. K. and P. Thakran (2018). "Intron specificity in pre-mRNA splicing." Curr Genet.

Mondal, M., M. Sengupta and K. Ray (2016). "Functional assessment of tyrosinase variants identified in individuals with albinism is essential for unequivocal determination of genotype-to-phenotype correlation." Br J Dermatol **175**(6): 1232-1242.

Muniz, Y. C., L. B. Ferreira, C. T. Mendes-Junior, C. E. Wiezel and A. L. Simoes (2008). "Genomic ancestry in urban Afro-Brazilians." Ann Hum Biol **35**(1): 104-111.

Nan, H., P. Kraft, A. A. Qureshi, Q. Guo, C. Chen, S. E. Hankinson, F. B. Hu, G. Thomas, R. N. Hoover, S. Chanock, D. J. Hunter and J. Han (2009). "Genome-wide association study of tanning phenotype in a population of European ancestry." J Invest Dermatol **129**(9): 2250-2257.

Newton, J. M., O. Cohen-Barak, N. Hagiwara, J. M. Gardner, M. T. Davisson, R. A. King and M. H. Brilliant (2001). "Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4." Am J Hum Genet **69**(5): 981-988.

Oetting, W. S. (2000). "The tyrosinase gene and oculocutaneous albinism type 1 (OCA1): A model for understanding the molecular biology of melanin formation." Pigment Cell Res **13**(5): 320-325.

Orlow, S. J. (1995). "Melanosomes are specialized members of the lysosomal lineage of organelles." J Invest Dermatol **105**(1): 3-7.

Park, J., A. H. Talukder, S. A. Lim, K. Kim, K. Pan, B. Melendez, S. D. Bradley, K. R. Jackson, J. S. Khalili, J. Wang, C. Creasy, B. F. Pan, S. E. Woodman, C. Bernatchez, D. Hawke, P. Hwu, K. M. Lee, J. Roszik, G. Lizee and C. Yee (2017). "SLC45A2: A Melanoma Antigen with High Tumor Selectivity and Reduced Potential for Autoimmune Toxicity." Cancer Immunol Res **5**(8): 618-629.

Paterson, E. K., T. J. Fielder, G. R. MacGregor, S. Ito, K. Wakamatsu, D. L. Gillen, V. Eby, R. E. Boissy and A. K. Ganesan (2015). "Tyrosinase Depletion Prevents the Maturation of Melanosomes in the Mouse Hair Follicle." PLoS One **10**(11): e0143702.

Pena, S. D., G. Di Pietro, M. Fuchshuber-Moraes, J. P. Genro, M. H. Hutz, S. Kehdy Fde, F. Kohlrausch, L. A. Magno, R. C. Montenegro, M. O. Moraes, M. E. de Moraes, M. R. de Moraes, E. B. Ojopi, J. A. Perini, C. Racciopi, A. K. Ribeiro-Dos-Santos, F. Rios-Santos, M. A. Romano-Silva, V. A. Sortica and G. Suarez-Kurtz (2011). "The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected." PLoS One **6**(2): e17063.

Pike, J. W. and M. B. Meyer (2010). "The vitamin D receptor: new paradigms for the regulation of gene expression by 1,25-dihydroxyvitamin D(3)." Endocrinol Metab Clin North Am **39**(2): 255-269, table of contents.

Pillaiyar, T., M. Manickam and S. H. Jung (2017). "Recent development of signaling pathways inhibitors of melanogenesis." Cell Signal **40**: 99-115.

Porto, I. O., C. T. Mendes-Junior, L. P. Felicio, R. C. Georg, P. Moreau, E. A. Donadi, J. A. Chies and E. C. Castelli (2015). "MicroRNAs targeting the immunomodulatory HLA-G gene: a new survey searching for microRNAs with potential to regulate HLA-G." Mol Immunol **65**(2): 230-241.

Ramsden, C. A. and P. A. Riley (2014). "Tyrosinase: the four oxidation states of the active site and their relevance to enzymatic activation, oxidation and inactivation." Bioorg Med Chem **22**(8): 2388-2395.

Rehmsmeier, M., P. Steffen, M. Hochsmann and R. Giegerich (2004). "Fast and effective prediction of microRNA/target duplexes." RNA **10**(10): 1507-1517.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Salzano, F. M. and M. C. Bortolini (2002). The Evolution and Genetics of Latin American Populations, Cambridge University Press.

Salzano, F. M. and M. Sans (2014). "Interethnic admixture and the evolution of Latin American populations." Genet Mol Biol **37**(1 Suppl): 151-170.

Santos, N. P., E. M. Ribeiro-Rodrigues, A. K. Ribeiro-Dos-Santos, R. Pereira, L. Gusmao, A. Amorim, J. F. Guerreiro, M. A. Zago, C. Matte, M. H. Hutz and S. E. Santos (2010). "Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel." Hum Mutat **31**(2): 184-190.

Santos, R. V. and C. E. A. Coimbra Jr. (1994). Saúde e povos indígenas. Rio de Janeiro, SciELO Books.

Shen, L. X., J. P. Babilion and V. P. Stanton, Jr. (1999). "Single-nucleotide polymorphisms can cause different structural folds of mRNA." Proc Natl Acad Sci U S A **96**(14): 7871-7876.

Siewierska-Gorska, A., A. Sitek, E. Zadzińska, G. Bartosz and D. Strapagiel (2017). "Association of five SNPs with human hair colour in the Polish population." Homo **68**(2): 134-144.

Simeonov, D. R., X. Wang, C. Wang, Y. Sergeev, M. Dolinska, M. Bower, R. Fischer, D. Winer, G. Dubrovsky, J. Z. Balog, M. Huizing, R. Hart, W. M. Zein, W. A. Gahl, B. P. Brooks and D. R. Adams (2013). "DNA variations in oculocutaneous albinism: an updated mutation list and current outstanding issues in molecular diagnostics." Hum Mutat **34**(6): 827-835.

Sitaram, A. and M. S. Marks (2012). "Mechanisms of protein delivery to melanosomes in pigment cells." Physiology (Bethesda) **27**(2): 85-99.

Smale, S. T. and J. T. Kadonaga (2003). "The RNA polymerase II core promoter." Annu Rev Biochem **72**: 449-479.

Solomon, E. I., U. M. Sundaram and T. E. Machonkin (1996). "Multicopper Oxidases and Oxygenases." Chem Rev **96**(7): 2563-2606.

Stephens, M., N. J. Smith and P. Donnelly (2001). "A new statistical method for haplotype reconstruction from population data." Am J Hum Genet **68**(4): 978-989.

Stokowski, R. P., P. V. Pant, T. Dadd, A. Fereday, D. A. Hinds, C. Jarman, W. Filsell, R. S. Ginger, M. R. Green, F. J. van der Ouderaa and D. R. Cox (2007). "A genomewide association study of skin pigmentation in a South Asian population." Am J Hum Genet **81**(6): 1119-1132.

Sturm, R. A. (2006). "A golden age of human pigmentation genetics." Trends Genet **22**(9): 464-468.

Sturm, R. A. (2009). "Molecular genetics of human pigmentation diversity." Hum Mol Genet **18**(R1): R9-17.

Sturm, R. A. and D. L. Duffy (2012). "Human pigmentation genes under environmental selection." Genome Biol **13**(9): 248.

Sulem, P., D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar, K. P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Palsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K. R.

Benediktsdottir, K. K. Aben, L. A. Kiemeney, J. H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir and K. Stefansson (2007). "Genetic determinants of hair, eye and skin pigmentation in Europeans." Nat Genet **39**(12): 1443-1452.

Sun, X. and L. E. Maquat (2000). "mRNA surveillance in mammalian cells: the relationship between introns and translation termination." RNA **6**(1): 1-8.

Talebzadeh, M. and F. Zare-Mirakabad (2014). "Transcription factor binding sites prediction based on modified nucleosomes." PLoS One **9**(2): e89226.

Theos, A. C., D. Tenza, J. A. Martina, I. Hurbain, A. A. Peden, E. V. Sviderskaya, A. Stewart, M. S. Robinson, D. C. Bennett, D. F. Cutler, J. S. Bonifacino, M. S. Marks and G. Raposo (2005). "Functions of adaptor protein (AP)-3 and AP-1 in tyrosinase sorting from endosomes to melanosomes." Mol Biol Cell **16**(11): 5356-5372.

Thorvaldsdottir, H., J. T. Robinson and J. P. Mesirov (2013). "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." Brief Bioinform **14**(2): 178-192.

Toth, L., B. Fabos, K. Farkas, A. Sulak, K. Tripolszki, M. Szell and N. Nagy (2017). "Identification of two novel mutations in the SLC45A2 gene in a Hungarian pedigree affected by unusual OCA type 4." BMC Med Genet **18**(1): 27.

Tully, G. (2007). "Genotype versus phenotype: human pigmentation." Forensic Sci Int Genet **1**(2): 105-110.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel and M. A. DePristo (2013). "From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline." Curr Protoc Bioinformatics **43**: 11 10 11-33.

Walsh, S., F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki and M. Kayser (2013). "The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA." Forensic Sci Int Genet **7**(1): 98-115.

Wang, H. F., L. Feng and D. K. Niu (2007). "Relationship between mRNA stability and intron presence." Biochem Biophys Res Commun **354**(1): 203-208.

Wasmeier, C., A. N. Hume, G. Bolasco and M. C. Seabra (2008). "Melanosomes at a glance." J Cell Sci **121**(Pt 24): 3995-3999.

Wei, A. H., D. J. Zang, Z. Zhang, X. M. Yang and W. Li (2015). "Prenatal genotyping of four common oculocutaneous albinism genes in 51 Chinese families." J Genet Genomics **42**(6): 279-286.

Wilde, S., A. Timpson, K. Kirsanow, E. Kaiser, M. Kayser, M. Unterlander, N. Hollfelder, I. D. Potekhina, W. Schier, M. G. Thomas and J. Burger (2014). "Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y." Proc Natl Acad Sci U S A **111**(13): 4832-4837.

Yamaguchi, Y., M. Brenner and V. J. Hearing (2007). "The regulation of skin pigmentation." J Biol Chem **282**(38): 27557-27561.

Yamaguchi, Y. and V. J. Hearing (2014). "Melanocytes and their diseases." Cold Spring Harb Perspect Med **4**(5).

Yamamoto, H. and T. Takeuchi (1981). "Immunoelectron microscopic localization of tyrosinase in the mouse melanocyte." J Histochem Cytochem **29**(8): 953-958.

Yuasa, I., K. Umetsu, S. Harihara, A. Kido, A. Miyoshi, N. Saitou, B. Dashnyam, F. Jin, G. Lucotte, P. K. Chattopadhyay, L. Henke and J. Henke (2006). "Distribution of the F374 allele of the SLC45A2 (MATP) gene and founder-haplotype analysis." Ann Hum Genet **70**(Pt 6): 802-811.

Yuasa, I., K. Umetsu, G. Watanabe, H. Nakamura, M. Endoh and Y. Irizawa (2004). "MATP polymorphisms in Germans and Japanese: the L374F mutation as a population marker for Caucasoids." Int J Legal Med **118**(6): 364-366.

Yuen, A. W. and N. G. Jablonski (2010). "Vitamin D: in the evolution of human skin colour." Med Hypotheses **74**(1): 39-44.

Zhang, Y. M., X. J. Zhou, F. J. Cheng, Y. Y. Qi, P. Hou, M. H. Zhao and H. Zhang (2016). "Polymorphism rs3828903 within MICB Is Associated with Susceptibility to Systemic Lupus Erythematosus in a Northern Han Chinese Population." J Immunol Res **2016**: 1343760.

10. ANEXOS

10.1 ANEXO A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

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.

Comitê de Ética em Pesquisa da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - USP
Avenida Bandeirantes, 3900 - bloco 3 - sala 16 - 14040-901 - Ribeirão Preto - SP - Brasil
Fone: (16) 3602-4811 / Fax: (16) 3633-2660 / E-mail: coetp@ffclrp.usp.br

- 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.

Assinatura do pesquisador responsável

Eu, _____, RG nº _____,
residente na _____ nº _____,
cidade de _____, telefone _____
abaixo assinado, tendo recebido as informações acima e ciente dos meus direitos, concordo em participar da pesquisa por livre e espontânea vontade. Declaro ainda que recebi uma cópia deste termo de consentimento assinada pelo pesquisador responsável.

Tendo ciência do exposto acima, assino abaixo.

Ribeirão Preto, ____ de _____ de 20__.

Assinatura do doador

Comitê de Ética em Pesquisa da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - USP
Avenida Bandeirantes, 3900 - bloco 3 - sala 16 - 14040-901 - Ribeirão Preto - SP - Brasil
Fone: (16) 3602-4811 / Fax: (16) 3633-2660 / E-mail: coetp@ffclrp.usp.br

10.2 ANEXO B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA ARMAZENAMENTO DE MATERIAL BIOLÓGICO

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.

Comitê de Ética em Pesquisa da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - USP
Avenida Bandeirantes, 3900 - bloco 3 - sala 16 - 14040-901 - Ribeirão Preto - SP - Brasil
Fone: (16) 3602-4811 / Fax: (16) 3633-2660 / E-mail: coetp@ffclrp.usp.br

- 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 _____, telefone _____
abaixo assinado, tendo recebido as informações acima e ciente dos 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 _____ de 20 ____.

Assinatura do doador

Comitê de Ética em Pesquisa da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - USP
Avenida Bandeirantes, 3900 - bloco 3 - sala 16 - 14040-901 - Ribeirão Preto - SP - Brasil
Fone: (16) 3602-4811 / Fax: (16) 3633-2660 / E-mail: coetp@ffclrp.usp.br

10.3 ANEXO C – APROVAÇÃO COMITÊ DE ÉTICA



Universidade de São Paulo
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Comitê de Ética em Pesquisa

Campus de Ribeirão Preto

Of.CEtP/FFCLRP-USP/028-vjmc

Ribeirão Preto, 10 de março de 2014

Prezado Pesquisador,

Comunicamos a V. Sa. que o projeto de pesquisa intitulado "SEQUENCIAMENTO DE NOVA GERAÇÃO DE GENES ENVOLVIDOS NA BIOSÍNTESE DE MELANINA EM AMOSTRA DA POPULAÇÃO BRASILEIRA" foi analisado pelo Comitê de Ética em Pesquisa da FFCLRP-USP, em sua 129ª Reunião Extraordinária, realizada em 06.03.2014, e enquadrado na categoria: **APROVADO** (CAAE n.º 25696413.7.0000.5407).

Solicitamos que eventuais modificações ou emendas ao projeto de pesquisa sejam apresentadas ao CEP, de forma sucinta, identificando a parte do projeto a ser modificada e suas justificativas, e que relatórios parciais e final sejam entregues, via Plataforma Brasil, conforme os itens X.1.3.d e XI.2.d, da Res. CNS nº 466/2012.

Atenciosamente,


Prof.ª Dr.ª ANDRÉIA SCHMIDT
Coordenadora

Ao Senhor
Prof. Dr. Celso Teixeira Mendes Junior
Departamento de Química da FFCLRP USP

CEP - Comitê de Ética em Pesquisa da FFCLRP USP
Fone: (16) 3602-4811
Fax: (16) 3633-2660 (direto) ou 3633-5015
Avenida Bandeirantes, 3900 - bloco 3 - sala 16 - 14040-901 - Ribeirão Preto - SP - Brasil
Homepage: <http://www.ffclrp.usp.br> - e-mail: coetp@ffclrp.usp.br

10.4 ANEXO D – 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	104
2. Quantificação e verificação da qualidade do DNA	104
3. Preparo dos reagentes para enriquecimento de alvo	105
4. Enriquecimento do alvo e preparo das biblioteca	105
4.1. Digestão e hibridação do DNA	106
4.2. Captura, ligação, eluição e PCR	107
4.3. Purificação da biblioteca	108
5. Controle de qualidade e quantificação da biblioteca	108
6. Sequenciamento	109
7. Referências	110

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™ 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 íntrons flanqueadores dos éxons 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 biotinizadas 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 476 μ L de *RE buffer* e 11,9 μ L de BSA e esse conteúdo é distribuído em 8 tubos. Em cada um dos 8 tubos é adicionado um par de enzimas de restrição (7 μ L de cada). Cada amostra é digerida pelos 8 pares de enzima separadamente. Para isso, cada amostra é distribuída em 8 tubos (5 μ L por tubo) onde são adicionados 5 μ L 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.

Hibridação: Para o mix de hibridação são misturados 650 μ L da *Hybridization Solution* e 260 μ L das sondas *Haloplex* em um tubo e 70 μ L dessa solução é distribuída por amostra. Adicionalmente, são diluídos nessa solução 10 μ L 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 (80 μ L somando todos os 8 tubos de digestão) é então transferido para os microtubos contendo o mix de hibridação (70 μ L) e seus índices correspondentes (10 μ L). 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

Captura: A captura dos fragmentos hibridados é feita com o uso de 520 μ L de *beads* magnéticas recobertas por estreptavidina. Com o auxílio de uma placa magnética (*DynaMag™-2 Magnet, Life Technologies*), o sobrenadante da solução de *beads* é removido e substituído por 520 μ L de *Capture Solution*. Deste conteúdo, 40 μ L é adicionado a cada solução de DNA hibridizado (160 μ L) 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 100 μ L 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.

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

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

PCR: Com o auxílio da placa magnética, coletamos o DNA eluído num total de 20 μ L por amostra e o transferimos para tubos contendo mix de PCR. Tal mix para a reação de PCR é preparado com 209,3 μ L de água MilliQ, 130 μ L de 5X *Herculase II Reaction Buffer*, 5,2 μ L de dNTPs (100mM, 25mM para cada dNTP), 13 μ L de cada um dos *primers forward* e *reverse* (25 μ M), 6,5 μ L de ácido acético (2M) e 13 μ L da polimerase *Herculase II Fusion*, somando um total de 30 μ L 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/µL 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™ dsDNA BR Assay*, *Life Technologies*) para composição adequada do *pool* de bibliotecas para o sequenciamento.

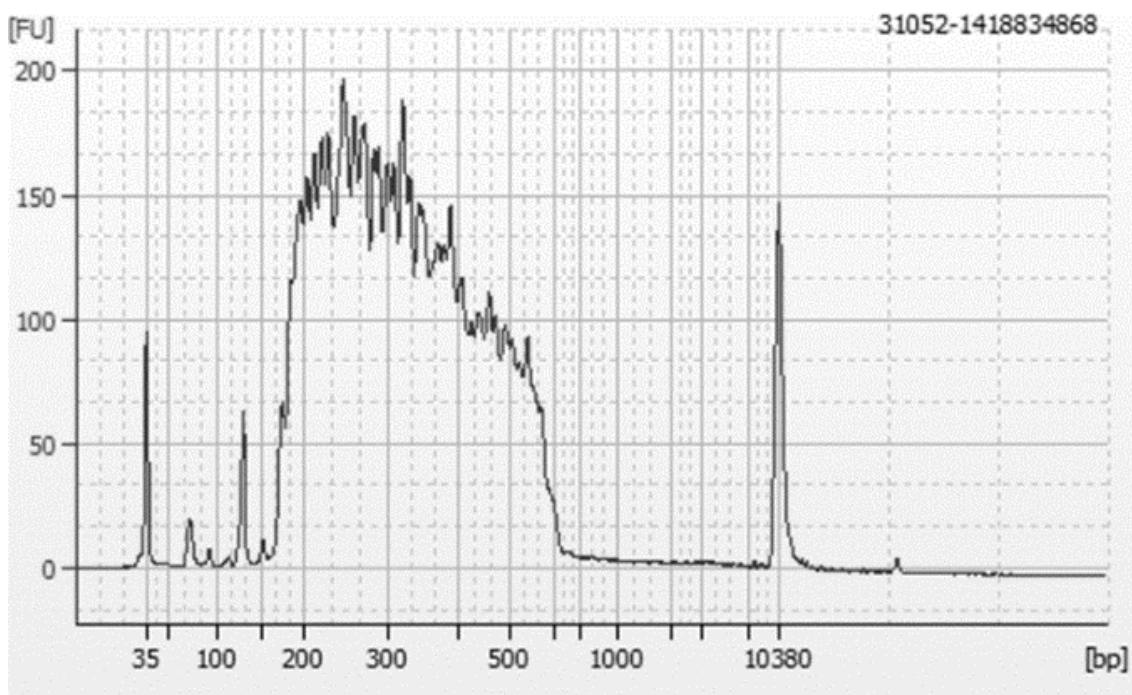


Figura 8: Perfil de picos esperado no eletroferograma para as bibliotecas de fragmento deste ensaio. Tal eletroferograma é produzido pela *Agilent Technologies* durante a avaliação do kit *Haloplex* customizado por meio de ensaio *in silico* de desenho de sondas para captura de DNA alvo utilizando o 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 superfície 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>

10.5 ANEXO E – PROTOCOLO DE PROCESSAMENTO COMPUTACIONAL DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO



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

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 – abril de 2018

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	112
1.1. Análise primária	112
1.2. Análise secundária	113
1.2.1. Avaliação da qualidade	113
1.2.2. Trimagem	113
1.2.3. Alinhamento ao genoma de referência	115
1.2.3.1. SAM	117
1.2.3.2. BAM	118
1.3. Análise terciária	119
1.3.1. Determinação de variantes	119
1.3.1.1. VCF	123
1.3.2. Remoção de genótipos de baixa qualidade	125
2. REFERÊNCIAS	127

1. ANÁLISE DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

O fluxograma de análises no NGS pode ser resumido 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* (<https://basespace.illumina.com>), no formato FastQ (*.fastq.gz), o qual possui uma pontuação de qualidade associada (*Q-score*) 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 $Q(X) = -10 \log_{10} (P(X))$ 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.

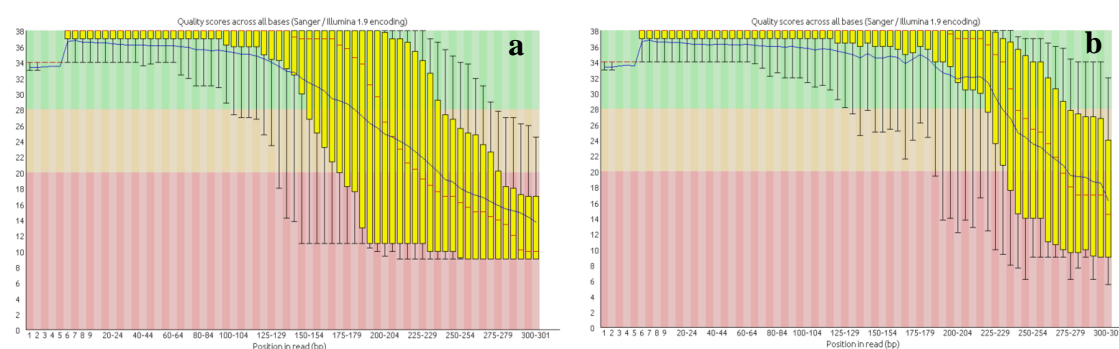


Figura 1: Gráfico de distribuição da qualidade por base no programa FastQC. Cada coluna apresenta um *boxplot* com os valores de qualidade média das bases incorporadas em todas as *reads* em um determinado ciclo. Como pode ser observado em (a) ocorre uma diminuição da qualidade ($Q < 30$) das *reads* 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, conseqüentemente, 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
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -o
sample.trimmed.r1.fastq -p sample.trimmed.r2.fastq
input.raw.r1.fastq input.raw.r2.fastq
```

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

Comandos utilizados no <i>cutadapt</i>	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> .
-o	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.

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: https://support.illumina.com/sequencing/sequencing_software/igenome.html. 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	<i>number of threads</i> (número de processos com um endereço/espaco compartilhado no <i>Linux</i>).
-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.

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

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.

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* (<http://samtools.sourceforge.net/>), 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 FIG0001 representa o nome da amostra e *sorted* indica que o arquivo BAM gerado se encontra ordenado:

```
FIG0001.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:

```
FIG0001.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*, ao encontrar uma região apresentando sinais de variação (região ativa), 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 restritivos 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 permite a adição posterior 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 \
  --emitRefConfidence 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, <i>HaplotypeCaller</i> ou <i>UnifiedGenotyper</i>
-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 \geq limiar mínimo são

	consideradas. O valor default recomendado é de 30
--	---

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 <i>HaplotypeCaller</i>
-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
-o	indica o <i>output</i> , que é um arquivo VCF ou gVCF (<i>Variant Call Format</i>) contendo os sítios que o algoritmo (<i>UnifiedGenotyper</i> ou <i>HaplotypeCaller</i> , respectivamente) identificou como sendo variáveis
--dontUseSoftClippedBases	direciona o programa a não analisar bases <i>soft 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 (<i>DuplicateRead</i> neste caso).

Quando o *HaplotypeCaller* é utilizado, após gerar os arquivos g.vcf para cada amostra, um passo adicional 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

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, <i>GenotypeGVCFs</i>
-I	indica arquivo de texto com lista de amostras (formato BAM) a ser utilizado pelo GATK

-o	indica o <i>output</i> , que é um arquivo gVCF (<i>Variant Call Format</i>) contendo os sítios que o algoritmo (<i>HaplotypeCaller</i>) 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):

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

Informações presentes no cabeçalho	Descrição das abreviações apresentadas
Fileformat	versão do formato de arquivo e a versão do <i>variant caller</i> utilizado
FILTER FORMAT e INFO	identifica quais filtros são aplicados aos dados essas linhas definem as anotações contidas nas 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 caller</i> , incluindo os nomes do arquivo com o 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

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 arquivo	Descrição das abreviações apresentadas
CHROM e POS	contém as coordenadas genômicas nas quais as variantes ocorrem
ID	campo opcional, cujo conteúdo se baseia nas informações contidas em um banco de dados de referência como o dbSNP
REF e ALT	identificam o alelo referência e alternativo(s) 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 ALT é representado pela base anterior à deleção
QUAL	é a probabilidade de que um polimorfismo (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	esse campo contém os filtros aplicados aos dados analisados. Se a variante passou 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 é precisa
CIPOS/CIEND	intervalo de confiança em torno da posição e posição final para variantes localizadas com baixa precisão
FORMAT	fornece a ordem das informações contidas nas colunas seguintes (amostras):
GT	indica o genótipo 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 <i>reads</i> 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 <i>ReadBackedPhasing</i>
PQ	qualidade de faseamento de haplótipos eventualmente determinados pelo <i>ReadBackedPhasing</i>

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, caso apenas a ferramenta *HaplotypeCaller* com o parâmetro `--emitRefConfidence GVCF` tenha sido previamente 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.

Por outro lado, caso ambas as ferramentas (*HaplotypeCaller* e *UnifiedGenotyper*) tenham sido previamente utilizadas, as variantes que apresentaram genótipos distintos nos dois arquivos VCF obtidos através do *HaplotypeCaller* e

UnifiedGenotyper, são analisadas manualmente após a análise pelo VCFx 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.

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. *Human Immunology*, 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 next-generation 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., Velì, 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>

10.6 ANEXO F – PROTOCOLO DE ANÁLISE ESTATÍSTICA DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO



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

Protocolo de análise estatística 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 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. FASEAMENTO EXPERIMENTAL	130
2. FASEAMENTO COMPUTACIONAL	131
2.1. Faseamento completo	132
2.2. Faseamento fragmentado	133
3. DESEQUILÍBRIO DE LIGAÇÃO	134
4. FREQUÊNCIAS ALÉLICAS, EQUILÍBRIO DE HARDY-WEINBERG E HETEROZIGOSE ESPERADA	135
5. ASSOCIAÇÕES GENÓTIPO-FENÓTIPO	135
6. REFERÊNCIAS	136

1. FASEAMENTO EXPERIMENTAL

É utilizada a ferramenta *ReadBackedPhasing*, também incorporada no programa GATK (DePristo et al., 2011; McKenna et al., 2010; Van der Auwera et al., 2013), a qual se utiliza da informação da leitura das *reads paired-end* para determinar a fase de alelos de variantes presentes em um mesmo par de *reads*.

A linha de comando a seguir é utilizada para realizar esse pré-faseamento (Tabela 1):

```
java -jar GenomeAnalysisTK.jar \  
  -T ReadBackedPhasing \  
  -R reference.fasta \  
  -I reads.bam \  
  --variant SNPs.vcf \  
  -L SNPs.vcf \  
  -o phased_SNPs.vcf \  
  --phaseQualityThresh 500.0
```

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

Comandos utilizados no <i>ReadBackedPhasing</i>	Função do comando executado
-T	indica qual ferramenta do GATK será utilizada. Neste caso, o <i>ReadBackedPhasing</i>
-R	indica o arquivo com o genoma de referência
-I	indica arquivo de texto com lista de amostras (formato BAM) a ser utilizado pelo <i>ReadBackedPhasing</i>
--variant	arquivo VCF consenso (*)
-L	direciona o GATK a restringir o processamento dos dados às variantes especificadas no arquivo “ <i>SNPs.vcf</i> ”
-o	indica o <i>output</i> , que é um arquivo VCF processado contendo os sítios de variação originais, com fase determinada para alelos de sítios próximos
--phaseQualityThresh	valor que indica a qualidade mínima de faseamento a ser empregada pelo algoritmo para aceitar uma fase como verdadeira. O valor <i>default</i> recomendado é 30. O valor de 500 aqui utilizado assegura que apenas fases bem definidas pelo GATK sejam aceitas como verdadeira

(*) Arquivo gerado a partir da comparação dos arquivos gerados pelo *HaplotypeCaller* e *UnifiedGenotyper*; posteriormente processados pelo VCFx.0.10.1 (Castelli et al., 2015). Vide Protocolo de processamento computacional de dados de sequenciamento de nova geração.

No entanto, a informação obtida através dos dados de sequenciamento ainda não é suficiente para determinar a fase completa dos haplótipos. Por esse motivo, o resultado de faseamento obtido através do *ReadBackedPhasing* deve ser incorporado ao método de faseamento computacional PHASE 2.1 (M Stephens, Smith, & Donnelly, 2001; Matthew Stephens & Donnelly, 2003).

2. FASEAMENTO COMPUTACIONAL

A partir das diversas configurações de haplótipo possíveis para os genótipos identificados são estimadas as probabilidades de uma determinada configuração haplotípica. Dentre as abordagens estatísticas disponíveis para o faseamento de indivíduos não aparentados, optamos pela utilização do programa PHASE 2.1 (M Stephens et al., 2001; Matthew Stephens & Donnelly, 2003).

O PHASE implementa um método de estatística Bayesiana para reconstruir os haplótipos a partir dos dados de genotipagem da população (Matthew Stephens & Donnelly, 2003). O método empregado pelo programa se baseia na utilização de modelos coalescentes, segundo os quais os novos haplótipos seriam derivados de haplótipos antigos pelos processos de mutação e recombinação (o que ajuda a entender a grande semelhança observada entre os haplótipos) e no modelo de *Hidden Markov* (HMM), um processo probabilístico segundo o qual a distribuição dos estados futuros depende apenas do estado atual e não de estados prévios. São utilizados os seguintes parâmetros para as análises realizadas no PHASE: número de interações (1000), *thinning interval* (1), *burn-in value* (100), -X10 (o software faz uma última corrida utilizando 10x o número de interações especificado), -x10 (o software aplica o algoritmo 10 vezes e devolve o melhor resultado). Como muitos sítios de variação podem se mostrar não-polimórficos em uma dada amostra populacional, levando o PHASE a inferir haplótipos com valores muito baixos de probabilidade, os *singletons* (contagem alélica = 1) são excluídos desta análise.

A determinação da fase dos haplótipos pode ser feita por meio de duas abordagens: faseamento completo e/ou faseamento fragmentado. O faseamento completo geralmente é mais indicado para regiões menores, em que as informações obtidas pelo *ReadBackedPhasing* são melhor aproveitadas pelo PHASE devido à (a) ausência de lacunas entre as regiões cobertas pelo ensaio de sequenciamento e (b) menor distância entre as variantes encontradas na região. Já o faseamento fragmentado é

mais indicado para regiões maiores, principalmente quando há (a) lacunas no sequenciamento (introns, por exemplo) e (b) as variantes estão distantes umas das outras (diferentes éxons, por exemplo), impossibilitando o faseamento experimental pelo *ReadBackedPhasing* e tornando o faseamento computacional mais complexo. Nesse caso, o faseamento pode ser feito para diferentes regiões (por exemplo, promotora, 5'UTR, CDS, 3'UTR) separadamente e posteriormente é feita uma etapa final de faseamento unindo os haplótipos encontrados.

2.1. Faseamento completo

Uma primeira corrida é realizada utilizando o arquivo VCF final obtido à partir da comparação entre os arquivos gerados pelos algoritmos *HaplotypeCaller* e *UnifiedGenotyper*. Adicionalmente, foram desenvolvidos *scripts* (Castelli E.C.) para implementar uma metodologia que faz uso das informações de fase definidas pelo algoritmo *ReadBackedPhasing*, utilizando a opção $-k$ do *software* PHASE. Nesta metodologia, para cada amostra são criados blocos de haplótipos de fase experimentalmente determinada. Em seguida, o programa PHASE realiza análises de inferência de haplótipos fixando cada um desses blocos e inferindo o restante do haplótipo de cada amostra, sendo o número de análises equivalente ao número máximo de blocos observado para as amostras que compõem o arquivo *input*. Além disso, o programa ainda faz uma última análise de inferências desconsiderando quaisquer informações do faseamento experimental. Com base nos resultados de haplótipos obtidos é realizada a maturação visual dos resultados, sendo selecionados indivíduos que tiveram o mesmo par de haplótipos inferido com probabilidade de pelo menos 80% em todas as corridas do programa.

Feita essa seleção, é realizada uma segunda corrida no PHASE especificando a fase completa dos indivíduos que tiveram seus haplótipos inferidos na corrida anterior (probabilidade $> 80\%$), permitindo com isso que o programa melhorasse seu desempenho na inferência dos haplótipos dos indivíduos restantes. Para tanto, o arquivo contendo as fases conhecidas (.known) é editado, onde as informações de fase daqueles indivíduos selecionados anteriormente são substituídas por "0", o que indica que a fase para todos os *loci* de determinado indivíduo é a mesma que a dos genótipos apresentados no arquivo *input* (.inp), o qual também é editado para conter a fase correta dos haplótipos desses indivíduos.

Após a determinação dos haplótipos, o arquivo *out.pairs* editado (contendo somente o par de haplótipos inferido com maior probabilidade para cada indivíduo) é utilizado como *input* para rodar o *script phase_to_vcf.pl* (Castelli E.C.), que gera como *output* um arquivo VCF faseado. Neste momento, as linhas com os *singletons* que tiveram a fase informada pelo *ReadBackedPhasing* podem ser reinseridas no arquivo VCF faseado, caso o pesquisador esteja certo de que aquela região em que o *singleton* se encontra tenha sido adequadamente sequenciada em todas as amostras.

2.2. Faseamento fragmentado

Primeiramente é preciso decidir em quantos fragmentos a região cromossômica será dividida. Essa decisão deve ser tomada levando em consideração a quantidade de variantes presentes em cada região e a existência de lacunas entre regiões cobertas (por exemplo, promotora, 5'UTR, éxons, 3'UTR).

Após essa separação, cada região deve ser faseada utilizando o software PHASE com os parâmetros definidos anteriormente. Os *scripts* citados no faseamento completo desenvolvidos por Castelli E.C. podem ser implementados nesse momento caso, para essa região, o *ReadBackedPhasing* tenha sido informativo. Terminado o faseamento, são aceitos como corretos os haplótipos inferidos com probabilidade superior a 80%, sendo os pares de haplótipos dos demais indivíduos considerados como *missing regions*.

Após ter determinado os haplótipos para cada fragmento, para definir o haplótipo estendido (completo) é preciso construir um novo arquivo de *input* para o PHASE em que cada fragmento faseado anteriormente é considerado como um *locus* polimórfico multi-alélico (M), e cada haplótipo dentro desse fragmento deve ser representado por um número de 1 a n , sendo n o número de haplótipos distintos encontrados neste fragmento. É importante enfatizar que um máximo de 49 haplótipos + *missing regions* (-1) pode estar presente em um determinado fragmento. Para atribuir a numeração aos haplótipos de cada fragmento (e eventualmente atribuir uma designação hierárquica definitiva, composta por linhagens e haplótipos derivados), é interessante avaliar a relação molecular entre os haplótipos e suas frequências por meio de uma *network*. Para isso podem ser usados os *softwares* Network (www.fluxus-engineering.com) ou PopART (<http://popart.otago.ac.nz>), por exemplo.

A inferência dos haplótipos estendidos deve ser feita utilizando os mesmos parâmetros citados anteriormente. Terminada a análise, devem ser aceitos como corretos os haplótipos inferidos com probabilidade superior a 80%. Nesse caso é importante verificar se os haplótipos estendidos inferidos em indivíduos que apresentavam *missing regions* em algum dos fragmentos são compatíveis com os genótipos originalmente determinados em tais regiões: se for inferido um haplótipo que envolve alelos inexistentes na *missing region* desse indivíduo, a inferência deve ser descartada.

3. DESEQUILÍBRIO DE LIGAÇÃO

O arquivo VCF deve ser convertido para o formato ARP através de *script* ferramenta VCFx (versão 0.10.1), disponível em http://www.castelli-lab.net/apps/apps_vcfx.php.

O desequilíbrio de ligação entre os *loci* estudados pode ser avaliado por meio do teste de razão de verossimilhança (Excoffier & Slatkin, 1998) implementado no programa Arlequin 3.5 (Excoffier & Lischer, 2010). Após a reconstrução dos haplótipos, com a fase gamética conhecida, o programa permite estimar os coeficientes de desequilíbrio de ligação entre pares de *loci* (D , D' e r^2).

O coeficiente clássico de desequilíbrio de ligação (D) mede o desvio da associação aleatória entre alelos nos diferentes *loci* (Lewontin & Kojima, 1960), sendo expresso por $D = p_{ij} - p_i p_j$, onde p_{ij} é a frequência do haplótipo formado pelos alelos i no primeiro *locus* e j no segundo e $p_i p_j$ corresponde ao produto das frequências dos alelos i e j .

Já (D') é o coeficiente de desequilíbrio de ligação (D) padronizado pelo máximo valor que esse pode assumir (D_{max}) de acordo com as frequências alélicas (Lewontin, 1964):

$$D'_{ij} = \frac{D_{ij}}{D_{ij\ max}}$$

Uma outra medida do desequilíbrio de ligação é representada pelo quadrado do coeficiente de correlação entre as frequências alélicas (r^2), que é expresso em função da medida D :

$$r^2 = \frac{D^2}{p_i(1-p_i) \cdot p_j(1-p_j)}$$

Para visualização gráfica dos blocos haplotípicos presentes ao longo das regiões analisadas e do desequilíbrio de ligação associado pode ser utilizado o programa *Haploview* 4.2 (Barrett, Fry, Maller, & Daly, 2005). Através dessa representação é possível então avaliar o nível de desequilíbrio presente entre cada par de SNP do segmento avaliado.

Os pares de *loci* são considerados em forte DL se o limite superior do IC95% de D' for igual ou superior a 0,98 (o que é consistente com ausência de recombinação histórica) e o limite inferior do intervalo for maior que 0,7. Assim, um bloco haplotípico é definido como uma região onde uma pequena parcela das comparações entre os pares de sítios bialélicos (<5%) apresenta forte evidência de recombinação histórica. Estes 5% são tolerados uma vez que outras forças evolutivas (como mutação recorrente, conversão gênica, erros de montagem do genoma ou de genotipagem), além da recombinação, podem romper os padrões de haplótipo. O método então ordena por tamanho a lista dos possíveis blocos de haplótipos, sendo cada bloco adicionado de forma a não haver sobreposição com o bloco anterior (Gabriel, 2002).

4. FREQUÊNCIAS ALÉLICAS, EQUILÍBRIO DE HARDY-WEINBERG E HETEROZIGOSE ESPERADA

As frequências alélicas para cada sítio de variação (tanto estratificadas por categoria fenotípica quanto globalmente) e a aderência das frequências genotípicas ao modelo de Equilíbrio de Hardy-Weinberg (EHW) podem ser verificadas através do programa GENEPOP v.4.0.10 (Raymond & Rousset, 1995; Rousset, 2008). A heterozigose esperada pode ser calculada a partir do programa Arlequin 3.5 (Excoffier & Lischer, 2010).

5. ASSOCIAÇÕES GENÓTIPO-FENÓTIPO

O teste exato de Fisher bi-caudal, implementado em planilha desenvolvida para esta finalidade (Mendes-Junior C.T.), a qual se utiliza de suplemento disponível em <http://www.obertfamily.com/software/fisherexact.html>, pode ser utilizado para avaliar a existência de associação dos genótipos/alelos/haplótipos com determinado fenótipo.

A mesma planilha realiza ainda o cálculo de *Odds Ratio* (OR) e do intervalo de confiança associado (IC) (Le, 2003; Van Belle, Fisher, Heagerty, & Lumley, 2004). Deve ser adotado o IC de 95% e o limite de significância deve ser estabelecido em $\alpha = 0,05$. O valor de OR permite verificar quantas vezes o caráter em questão (fenótipo avaliado) é mais frequente no grupo de indivíduos apresentando determinado genótipo/alelo/haplótipo, comparativamente ao risco em indivíduos não portador deste fator (Rumel, 1986).

6. REFERÊNCIAS

- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–265. <https://doi.org/10.1093/bioinformatics/bth457>
- 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. *Human Immunology*, *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 next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–8. <https://doi.org/10.1038/ng.806>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Excoffier, L., & Slatkin, M. (1998). Incorporating genotypes of relatives into a test of linkage disequilibrium. *American Journal of Human Genetics*, *62*(1), 171–180. <https://doi.org/10.1086/301674>

- Gabriel, S. B. (2002). The Structure of Haplotype Blocks in the Human Genome. *Science*, 296(5576), 2225–2229. <https://doi.org/10.1126/science.1069424>
- Le, C. T. (2003). *Introductory Biostatistics*. Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471308889>
- Lewontin, R. C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*, 49(1), 49–67. <https://doi.org/10.1038/nature08185>
- Lewontin, R. C., & Kojima, K. (1960). The Evolutionary Dynamics of Complex Polymorphisms. *Evolution*, 14(4), 458. <https://doi.org/10.2307/2405995>
- 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>
- Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, 86, 248–249.
- Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Rumel, D. (1986). “Odds Ratio”: Algumas Considerações. *Rev Saúde Pública*, 20(3), 253–258. <https://doi.org/10.1590/S0034-89101986000300011>
- Stephens, M., & Donnelly, P. (2003). Report A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *Am. J. Hum. Genet.*, 73(2002), 1162–1169. <https://doi.org/10.1086/379378>
- Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68(4), 978–989. <https://doi.org/10.1086/319501>
- Van Belle, G., Fisher, L. D., Heagerty, P. J., & Lumley, T. (2004). *Biostatistics: A Methodology for the Health Sciences*. Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471602396>

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>

10.7 ANEXO G – PROTOCOLO PARA IDENTIFICAÇÃO COMPUTACIONAL DE INTERAÇÕES ENTRE miRNAS E REGIÃO 3'UTR



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

Protocolo para identificação computacional de interações entre miRNAs e região 3'UTR

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

Fracasso NCA, Silva GV, Castelli EC, 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.

MicroRNAs (miRNAs) atuam na regulação de diversos processos biológicos através da interação entre sua região *seed* e sequências presentes principalmente na região codificadora e 3'UTR dos mRNAs alvo (Mione & Bosserhoff, 2015). Essa interação, dependendo da complementariedade de sequência entre o miRNA e seu alvo, pode resultar na inibição da tradução do mRNA (no caso de complementariedade incompleta) ou facilitar a degradação do mRNA (havendo complementariedade completa) (Huntzinger & Izaurralde, 2011).

Características específicas dos miRNAs (como possibilidade de pareamento incompleto e pequeno tamanho da sequência – 19-25 nucleotídeos) permitem que essas moléculas regulem centenas de mRNAs, tornando bastante complexo o entendimento desta interação (Di Leva & Croce, 2013). Além disso, é necessário considerar que esta regulação pode ainda ser influenciada por variações gênicas presentes nos sítios de mRNA alvo (Palmero et al., 2011). Assim, a compreensão do papel desempenhado por

essas moléculas perpassa a identificação dos seus alvos no genoma. Essa identificação pode ser realizada experimentalmente. Todavia, tendo em vista a complexidade e os custos associados ao método experimental, tem-se buscado alternativas computacionais *in silico* para o reconhecimento preditivo da ligação miRNA-mRNA-alvo (Marín & Vaníek, 2011).

Na predição computacional, utiliza-se as sequências de miRNAs disponíveis em bancos de dados para buscar possíveis alvos, neste caso, sequências 3'UTR do mRNA de determinado gene. A restrição da análise à 3'UTR se deve ao fato de a maioria dos miRNAs se ligarem preferencialmente nessa região (Lewis, Shih, Jones-Rhoades, Bartel, & Burge, 2003).

A fim de identificar miRNAs que se liguem a sequências polimórficas da 3'UTR de um determinado gene, são selecionados os haplótipos encontrados com frequência acima de 1% na amostra populacional.

Tendo em vista o grande número de falso-positivos obtidos através dos métodos *in silico*, optamos por utilizar três algoritmos distintos para predição de miRNAs com afinidade a sequências da 3'UTR: RNAhybrid (Rehmsmeier, Steffen, Höchsmann, Giegerich, & Ho, 2004), miRanda (John et al., 2004) e intaRNA (Busch, Richter, & Backofen, 2008).

Considerando que o miRNA se hibridize ao alvo de maneira energeticamente ótima, o RNAhybrid se baseia no cálculo da energia mínima livre (MFE) de hibridização entre as posições iniciais (*seed*) no miRNA e no alvo. O miRanda também se baseia no cálculo da MFE de hibridização para predição de miRNAs afins, mas utiliza ainda informações relacionadas à conservação evolutiva da sequência alvo e à posição do alvo em 3'UTRs alinhadas de genes homólogos. Por sua vez, o intaRNA se baseia em um *score* combinado da energia de interação, que é calculado como a soma da energia livre de hibridização e da energia livre necessária para tornar os sítios de interação acessíveis.

Para realizar uma avaliação conjunta utilizando esses três algoritmos é empregado um *software*, mirhunt (Porto et al., 2015) (http://www.castellilab.net/apps/apps_mirhunt.php), que automatiza a análise como brevemente descrito: a) as sequências de mRNA da 3'UTR são fragmentadas em uma série de subsequências de 500 nucleotídeos com uma sobreposição de 50 nucleotídeos iniciais e finais; b) os três

algoritmos são utilizados para prever interações miRNA/mRNA considerando cada uma das subsequências; c) os resultados gerados apresentam todas as interações miRNA/mRNA encontradas por cada algoritmo; d) mantem-se nos resultados somente a ligação mais estável nos casos em que são identificadas mais de uma interação para um mesmo miRNA com afinidade para uma mesma região-alvo (para reconhecimento de uma mesma região alvo, os últimos nucleotídeos pareados da região *seed* são utilizados como referência, com uma tolerância máxima de 3 nucleotídeos); e) os resultados encontrados pelos três algoritmos são comparados; f) um sistema de pontuação é aplicado para cada interação miRNA/alvo. Os parâmetros utilizados para o miRanda são: -4 para extensão e -9 para abertura do *gap*, limiar de pontuação estabelecido como 50 e mínima energia livre configurada para -1 kcal/mol. Para o RNAhybrid: mínima energia livre de ligação configurada para -1 kcal/mol e número de *hits* ajustado para 10. O intaRNA é um método livre de parâmetro, então a configuração padrão é utilizada (Porto et al., 2015).

O banco de dados contendo os miRNAs humanos até então identificados (2.588) pode ser obtido a partir do miRBase, release 21 de junho de 2014 (Griffiths-Jones, Grocock, van Dongen, Bateman, & Enright, 2006). Versão mais recente, caso disponível, deve ser utilizada.

Os algoritmos são utilizados para inferir a melhor interação miRNA/mRNA (menor valor de MFE) para cada miRNA e cada sequência da 3'UTR disponível para análise. Finalmente, um sistema de pontuação é aplicado para cada conjunto de dados, como descrito detalhadamente em Porto et al. (2015). Essas pontuações têm como objetivo identificar as interações mais estáveis e/ou específicas entre miRNAs e seus alvos, permitindo a seleção de miRNAs mais influentes, evitando o uso de limites arbitrários. Ao compor os resultados dos três métodos, interações falso-positivas eventualmente identificadas por apenas um dos métodos são descartadas por não resultarem em uma alta pontuação.

A primeira pontuação (denominada valor ϵ) reflete a estabilidade da interação miRNA/mRNA. O menor valor ϵ encontrado por cada algoritmo é utilizado como referência do algoritmo para propósito de normalização. Este procedimento tem como objetivo tornar os MFEs dos diferentes algoritmos comparáveis. A pontuação de ϵ para cada interação miRNA/mRNA é calculada a partir da razão entre o MFE daquela interação e aquele da referência. Após a normalização, a pontuação de energia composta

para cada interação miRNA/mRNA é obtida pela soma das pontuações de cada algoritmo: $\varepsilon = \varepsilon_{\text{RNAhybrid}} + \varepsilon_{\text{miRanda}} + \varepsilon_{\text{IntaRNA}}$. Assim, quanto maior for a pontuação ε maior a probabilidade de ocorrer uma forte interação entre o miRNA e o alvo.

A segunda pontuação (denominada valor s) reflete a especificidade da interação miRNA/mRNA, comparando as interações observadas com uma distribuição nula que leva em consideração a interação destes 2588 miRNAs em 23.464 sequências 3'UTR de genes humanos. Quanto maior for a pontuação s , maior a probabilidade do miRNA em questão interagir somente com o gene estudado.

A terceira pontuação (denominada valor εs) é calculada como um valor combinado que reflete ambas as pontuações, ε e s . Quanto maior a pontuação εs , maior a probabilidade de que o miRNA em questão interaja de forma forte e específica com o gene estudado.

REFERÊNCIAS

- Busch, A., Richter, A. S., & Backofen, R. (2008). IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24(24), 2849–2856. <https://doi.org/10.1093/bioinformatics/btn544>
- Di Leva, G., & Croce, C. M. (2013). MiRNA profiling of cancer. *Current Opinion in Genetics and Development*, 23(1), 3–11. <https://doi.org/10.1016/j.gde.2013.01.004>
- Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, 34(Database issue), D140-4. <https://doi.org/10.1093/nar/gkj112>
- Huntzinger, E., & Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews. Genetics*, 12(2), 99–110. <https://doi.org/10.1038/nrg2936>
- John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., & Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biology*, 2(11), e363. <https://doi.org/10.1371/journal.pbio.0020363>

- Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). 33-Prediction of Mammalian MicroRNA Targets. *Cell*, 115(7), 787–798. [https://doi.org/10.1016/S0092-8674\(03\)01018-3](https://doi.org/10.1016/S0092-8674(03)01018-3)
- Marín, R. M., & Vaníek, J. (2011). Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Research*, 39(1), 19–29. <https://doi.org/10.1093/nar/gkq768>
- Mione, M., & Bosserhoff, A. (2015). MicroRNAs in melanocyte and melanoma biology. *Pigment Cell and Melanoma Research*, 28(3), 340–354. <https://doi.org/10.1111/pcmr.12346>
- Palmero, E. I., de Campos, S. G. P., Campos, M., de Souza, N. C. N., Guerreiro, I. D. C., Carvalho, A. L., & Marques, M. M. C. (2011). Mechanisms and role of microRNA deregulation in cancer onset and progression. *Genetics and Molecular Biology*, 34(3), 363–370. <https://doi.org/10.1590/S1415-47572011000300001>
- Porto, I. O. P., Mendes-Junior, C. T., Felício, L. P., Georg, R. C., Moreau, P., Donadi, E. A., ... Castelli, E. C. (2015). MicroRNAs targeting the immunomodulatory HLA-G gene: A new survey searching for microRNAs with potential to regulate HLA-G. *Molecular Immunology*, 65(2), 230–241. <https://doi.org/10.1016/j.molimm.2015.01.030>
- Rehmsmeier, M., Steffen, P., Höchsmann, M., Giegerich, R., & Ho, M. (2004). Fast and effective prediction of microRNA / target duplexes. *Spring*, (2003), 1507–1517. <https://doi.org/10.1261/rna.5248604>

11. 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.

DIVERSIDADE GENÉTICA DAS REGIÕES REGULATÓRIAS E CODIFICANTES DOS GENES *SLC45A2* E *TYR* EM AMOSTRA DA POPULAÇÃO BRASILEIRA

Nádia Carolina de Aguiar Fracasso¹, Alison Luis Eburneo Pereira¹, Maria Luiza Guimarães de Oliveira¹, Guilherme Debortoli¹, Leticia Marcorin¹, Guilherme do Valle Silva², Juliana Doblaz Massaro³, Eduardo Antônio Donadi³, Aguinaldo Luiz Simões¹, Erick da Cruz Castelli⁴, Celso Teixeira Mendes-Junior².

¹Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto - SP, Brasil.

²Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto - SP, Brasil.

³Departamento de Clínica Médica, Divisão de Imunologia Clínica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto - SP, Brasil.

⁴Departamento de Patologia, Faculdade de Medicina, Universidade Estadual Paulista, 18618-687, Botucatu - 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

O gene *SLC45A2* codifica a proteína MATP, envolvida na síntese de melanina através do processamento e transporte intracelular da tirosinase e transporte de prótons para o melanossomo. Por sua vez, a proteína tirosinase, codificada pelo gene *TYR*, catalisa os dois primeiros passos da conversão de tirosina em melanina, além de atuar em uma etapa final da biossíntese de eumelanina. Considerando que polimorfismos nestes genes influenciam a variação de características normais de pigmentação, sua aplicação na predição fenotípica pode ser de grande relevância para prática forense. Dessa forma, o presente trabalho teve como objetivo identificar polimorfismos nos genes *SLC45A2* e *TYR* que contribuam para a determinação das características de pigmentação em uma amostra da população brasileira. As regiões regulatórias e codificadoras dos genes *SLC45A2* e *TYR* foram analisadas por sequenciamento de nova geração em uma amostra de 340 indivíduos, estratificados de acordo com a pigmentação dos olhos, cabelos e pele. Bibliotecas de DNA foram preparadas utilizando o *Haloplex Target Enrichment System (Agilent Technologies)* e sequenciadas por meio da plataforma *MiSeq (Illumina)*. Os softwares CutAdapt, BWA e GATK/VCFx foram utilizados para trimagem dos adaptadores, alinhamento e chamada de variantes, respectivamente. O programa PHASE foi utilizado para a reconstrução dos haplótipos. Um total de 58 variantes foram identificadas no gene *SLC45A2*. Destas, 28 foram associadas a pelo menos uma das características de pigmentação avaliadas. Embora as regiões não-codificadoras tenham apresentado maior diversidade genética, associações significativas também foram encontradas em regiões exônicas. Dentre estas, destaca-se a associação do alelo rs16891982*G (exon 5) com pele clara ($p = 9,54 \times 10^{-35}$, OR = 22,66) e cabelos loiros ($p = 1,59 \times 10^{-26}$, OR = 31,72). Quando observamos os haplótipos inferidos para as regiões codificantes das isoformas 1 e 2 que possuem esse SNP e as associações encontradas, podemos notar que os únicos haplótipos associados com fenótipos de pigmentação claros para ambas isoformas [“Iso2cds1”, pele clara ($p = 3,65 \times 10^{-29}$, OR = 22,63) e cabelos loiros ($p = 6,34 \times 10^{-23}$, OR = 26,19) e “Iso1cds1”, pele clara ($p = 1,08 \times 10^{-24}$, OR = 17,38) e cabelos loiros ($p = 1,94 \times 10^{-22}$, OR = 22,93)] possuem como diferença em relação a todos os outros haplótipos associados com pigmentação escura o alelo rs16891982*G. Para o gene *TYR* foram identificadas 42 variantes e 15 se mostraram associadas a algum dos fenótipos de pigmentação avaliados. A maior parte da diversidade desse gene foi encontrada na região intrônica, com ênfase para a associação do genótipo rs1393350*A/A com olhos azuis ($p = 0,0253$, OR = 13,06) e cabelos castanho-claros ($p = 0,0019$, OR = 16,07). Ao analisarmos as associações encontradas para os

haplótipos inferidos para essa região, podemos notar que os haplótipos “cds5” ($p = 1,71 \times 10^{-05}$, OR = 21,26), “cds7” ($p = 0,0061$, OR = 23,70) e “cds9” ($p = 0,0017$, OR = 29,25) estão associados a pele escura e o haplótipo “cds12” ($p = 0,015$, OR = 21,61) a ausência de sardas. Quando nos atentamos a composição desses haplótipos em relação ao SNP rs1393350, podemos perceber que todos os haplótipos possuem o alelo referência (rs1393350*G), em oposição ao genótipo (rs1393350*A/A) que determina fenótipos de pigmentação claros. Esses resultados evidenciam o papel desempenhado pelos genes *SLC45A2* e *TYR* na pigmentação humana e sugerem sua utilidade na predição fenotípica.

Palavras-chave: *SLC45A2*, *TYR*, SNPs, Pigmentação, População brasileira.

INTRODUÇÃO

As características da pigmentação humana, tais como variações na cor dos cabelos, olhos e pele, pertencem à categoria dos mais visíveis e diferenciáveis fenótipos humanos. A base genética por trás da variação dessas características tem sido objeto de intenso estudo dos pesquisadores de diferentes áreas do saber.

A pigmentação humana é primariamente explicada pela presença de melanina na epiderme, íris e cabelos e sua variação fenotípica é determinada pela quantidade, tipo e distribuição da melanina pelos melanócitos, células especializadas na síntese e distribuição deste biopolímero (Hearing 2011, Kawakami and Fisher 2011, Liu, Wen et al. 2013). Muitos genes distintos direta ou indiretamente relacionados à pigmentação já foram identificados. Vários deles afetam processos críticos ao desenvolvimento dos melanoblastos (células precursoras dos melanócitos), outros regulam a diferenciação e sobrevivência dos melanócitos, alguns estão envolvidos na biogênese e funcionamento dos melanossomos e outros regulam processos distintos que influenciam a pigmentação. Por exemplo, genes que codificam enzimas críticas incluem o gene da tirosinase (*TYR*), *TYRP1* e *DCT*, cujas mutações afetam drasticamente a quantidade e qualidade da melanina sintetizada. Genes de proteínas estruturais críticas incluem *Pmel17* e *MART1*, ambas necessárias para a maturação estrutural do melanossomo. Existem ainda, genes que codificam proteínas envolvidas no processo de sinalização (*MC1R*) e proteínas envolvidas na distribuição e transporte de proteínas/íons aos melanossomos (*SLC45A2* e *SLC24A5*) (Yamaguchi, Brenner et al. 2007).

O gene *SLC45A2* (*solute carrier family 45, member 2*) codifica a proteína MATP (*membrane associated transport protein*) que está envolvida na síntese de melanina através do transporte de prótons via membrana melanossomal e, conseqüentemente, controle do pH e da homeostase iônica dentro do melanossomo. Adicionalmente, atua também no transporte intracelular da tirosinase por meio de um mecanismo ainda não identificado (Tully 2007, Dooley, Schwarz et al. 2013, Kondo, Namiki et al. 2015). O gene *SLC45A2* está localizado na região cromossômica 5p13.2 e tem 7 éxons abrangendo pouco mais de 40 Kb (Graf, Hodgson et al. 2005). A proteína MATP se distribui ao longo da bicamada lipídica da membrana melanossomal, assim como a proteína P associada ao OCA2 (Costin, Valencia et al. 2003). Existe a hipótese de que a proteína P seja uma transportadora de ânions na membrana melanossomal. Se a MATP está similarmente localizada na membrana e transporta prótons para controle do pH melanossomal, então a ação dessas duas proteínas pode resultar numa acidificação normal do melanossomo (Newton, Cohen-Barak et al. 2001).

Como mencionado, a proteína MATP também está relacionada ao transporte de tirosinase (TYR) através de um mecanismo ainda desconhecido (Kondo, Namiki et al. 2015). A enzima TYR é responsável pela conversão da tirosina em L-DOPA no primeiro passo da via bioquímica da produção da melanina. Sob condições normais, a MATP aumenta o pH melanossomal funcionando como um transportador de prótons. Deste modo, o cobre pode se ligar a tirosinase e torná-la ativa. Quando mutada, MATP é responsável pelo albinismo oculocutâneo tipo 4 (OCA4) (Newton, Cohen-Barak et al. 2001, Kondo, Namiki et al. 2015). Nos melanossomos dos indivíduos afetados por OCA4, a MATP não funciona normalmente e o lúmen melanossomal se torna ácido. Assim, o cobre não consegue ser incorporado pela TYR, reduzindo a atividade da mesma (Bin, Bhin et al. 2015).

O gene *TYR* (*tyrosinase*) está localizado na região cromossômica 11q14.3. Este gene possui 117.888pb, 5 éxons e codifica a enzima tirosinase composta por 529 aminoácidos. Tal proteína catalisa os dois primeiros passos da conversão de tirosina em melanina, além de atuar em uma etapa final da biossíntese de eumelanina. Quando mutada é responsável pelo albinismo oculocutâneo tipo 1 (OCA1) (Bae-Harboe and Park 2012).

Apesar da tirosinase ser uma das oxidases multicobre mais bem estudadas, de sua atividade de conversão de tirosina em melanina ter sido descoberta a mais de um século e de mais de 200 mutações terem sido associadas com albinismo (Solomon, Sundaram et al. 1996, Oetting 2000, Mondal, Sengupta et al. 2016), muitas informações sobre sua estrutura e função, os processos bioquímicos que levam ao transporte adequado dessa enzima para os melanossomos e como estes fatores e a

diversidade de polimorfismos em seu gene influenciam a associação de alelos com a pigmentação normal ainda precisam ser elucidados (Ramsden and Riley 2014, Kondo, Namiki et al. 2015, Lai, Wichers et al. 2018). Apesar do conhecimento de que a proteína codificada pelo gene *SLC45A2* está envolvida no transporte e distribuição de tirosinase funcional para os melanossomos, o mecanismo pelo qual ambas proteínas interagem e como isso pode influenciar a determinação das cores de cabelos, olhos e pele em humanos ainda não foi descoberto (Kondo, Namiki et al. 2015), todavia, avaliar a diversidade genética desses dois genes (*SLC45A2* e *TYR*) em uma população miscigenada como a brasileira pode expandir nosso conhecimento sobre a pigmentação humana. Sendo assim, informações advindas das análises dos polimorfismos de genes como o *SLC45A2* e *TYR*, seriam de grande importância na medicina personalizada (distúrbios relacionados a hiper e hipopigmentação), genética forense e em estudos sobre adaptação, seleção e evolução humana.

MATERIAL E MÉTODOS

Amostra populacional e determinação fenotípica dos indivíduos

O projeto de pesquisa do presente trabalho, juntamente com o questionário sobre características étnicas e morfológicas foi aprovado pela Comissão de Ética em Pesquisa da Instituição (FFCLRP-USP) de acordo com o processo CAAE nº 25696413.7.0000.5407.

Foram coletadas amostras de sangue (10ml) de um total de 340 voluntários não aparentados dos sexos feminino e masculino, com idade variando de 18 a 70 anos, recrutados, no Hemocentro de Ribeirão Preto (HCFMRP-USP e Posto de Coleta - Centro). Por meio de questionário, os indivíduos foram classificados em grupos de acordo com a cor dos olhos, cabelos e pele, bem como quanto à presença ou ausência de efélides (sardas). Cada indivíduo participante assinou um Termo de Consentimento Livre e Esclarecido (TCLE) permitindo o uso de seu material biológico (amostra de sangue) para estudo. O material coletado foi armazenado em tubos rotulados de modo que o doador não pudesse ser diretamente identificado, mantendo, assim, o caráter confidencial da informação.

A classificação dos indivíduos de acordo com a pigmentação da pele (I a VI) foi realizada segundo o sistema proposto por Fitzpatrick (Fitzpatrick 1988). Adicionalmente, a intensidade da pigmentação da pele foi quantificada de forma objetiva por meio do espectrofotômetro portátil CM-2600d/2500d (*Konica Minolta Holdings, Inc*). A cor dos olhos foi caracterizada e dividida em categorias compreendendo azul, verde, mel, castanho-claro e castanho-escuro. A cor dos

cabelos, por sua vez, foi caracterizada e dividida em categorias compreendendo ruivo, loiro-claro, loiro-escuro, castanho-claro, castanho-escuro e preto.

Análise laboratorial e sequenciamento

A extração do DNA das amostras de sangue foi realizada através de modificação do protocolo de *salting-out* (Miller, Dykes et al. 1988). De forma concisa, o processo consiste em etapas consecutivas de lise das hemácias, precipitação de proteínas (que serão descartadas) e lise dos linfócitos para expor o DNA. Após extração do DNA, foi realizada a quantificação por eletroforese em gel de agarose, fluorescência (*Qubit – Life Technologies*) e medidas espectrofotométricas (*Thermo Scientific NanoDrop™ 1000 Spectrophotometer*). A eletroforese em gel de agarose permitiu avaliar o nível de degradação do DNA genômico. O *NanoDrop®* foi utilizado para estimar a concentração inicial do DNA e avaliar o grau de contaminação do material genômico por proteínas. Já o *Qubit™* foi utilizado para avaliar a concentração de DNA genômico por meio do *dsDNA BR Assay (Invitrogen, Carlsbad, CA)*, de modo que diluições permitissem que este ficasse a uma concentração de 5ng/μL, normalizando assim as concentrações para a utilização deste na montagem das bibliotecas. Após a quantificação, bibliotecas para cada amostra de DNA foram preparadas utilizando-se o *HaloPlex Target Enrichment System Kit*. Para a montagem de cada biblioteca são necessárias as seguintes etapas: digestão do DNA genômico por enzimas de restrição e hibridização dos fragmentos desejados com sondas biotinizadas cujas extremidades fossem complementares a esses fragmentos-alvo (durante esse passo os fragmentos são circularizados e índices, que irão identificar cada indivíduo durante o sequenciamento, são adicionados), captura desses fragmentos utilizando *beads* magnéticas recobertas por estreptavidina, ligação dos fragmentos circularizados, eluição e, finalmente, PCR dos fragmentos-alvo capturados. Após a amplificação, as bibliotecas de fragmentos geradas para cada indivíduo foram purificadas por meio de *beads* paramagnéticas utilizando-se o kit *Agencourt® AMPure® XP (Beckman Coulter, Inc)*. O controle de qualidade das bibliotecas e a normalização das mesmas foi realizado pelo sistema de eletroforese capilar *2100 Bioanalyzer (Agilent Technologies)* e por fluorescência (*Qubit™ dsDNA BR Assay, Life Technologies*). O sequenciamento foi realizado junto ao Departamento de Clínica Médica da FMRP-USP, por meio da plataforma *MiSeq Personal Sequencer (Illumina Inc.)* utilizando o *MiSeq Reagent Kit V3 (600 cycle)*, seguindo-se as recomendações do fabricante. O sequenciamento foi do tipo leitura pareada das duas extremidades de cada fragmento (*paired-end*), aumentando o tamanho das sequências obtidas e melhorando a qualidade do mapeamento dessas sequências à sequência de referência.

Análises de dados pós-sequenciamento e estatísticas

Após o sequenciamento, as etapas seguintes de análises podem ser sumarizadas em duas fases principais:

(1) a fase de pré-processamento e chamada de variantes, que consistiu na geração dos dados no formato FastQ, aplicação de filtros de qualidade, alinhamento a um genoma de referência e determinação das variantes através de comparação das sequências alvo a um genoma de referência. O *software* FastQC (Andrews 2010) foi empregado para gerar um relatório completo da qualidade das sequências presentes em um arquivo FastQ. A trimagem foi realizada com uso da ferramenta *cutadapt* (Martin 2011). O mapeamento contra um genoma de referência (*Homo sapiens*, UCSC, hg19), disponibilizado pela *Illumina* foi indexado por meio do índice FM (Ferragina, Manzini et al. 2004). Um pacote de softwares, denominado BWA, foi utilizado para realizar esse alinhamento. O BWA usa um algoritmo do tipo *Burrows-Wheeler* (BWT) que faz um processamento estatístico de reconstrução de um bloco de dados original de forma rápida e baixo consumo de memória (Li and Durbin 2009). O *output* originado é um arquivo SAM (*.sam) que foi convertido pelo programa *SAMtools* (<http://samtools.sourceforge.net/>) em um arquivo BAM (*.bam) porque este último é menor que o SAM e pode ser visualizado em programas como o IGV ou o *UCSC Genome Browser*. Após o alinhamento das *reads* ao genoma de referência, o programa *SAMStat* (Lassmann, Hayashizaki et al. 2011) foi utilizado para verificar a acurácia do mapeamento. O processo de determinação de variantes foi realizado por meio da comparação das regiões genômicas sequenciadas com um genoma de referência. O programa utilizado para tal análise foi o *Genome Analysis Toolkit* (GATK) v.3.5, disponibilizado pelo *Broad Institute* (McKenna, Hanna et al. 2010, DePristo, Banks et al. 2011, Van der Auwera, Carneiro et al. 2013). Para realizar a determinação de SNPs foi utilizada a ferramenta *HaplotypeCaller* do GATK. Utilizando o *HaplotypeCaller*, foi gerado um arquivo GVCF (*.g.vcf) com o objetivo de obter informações adicionais sobre os sítios de variação. Genótipos de baixa cobertura e alelos foram interrogados, submetendo o arquivo VCF obtido pelo *HaplotypeCaller* (GVCF) à rotina do *script* VCFx (Castelli, Mendes-Junior et al. 2015, Lima, Buttura et al. 2016).

(2) a fase final de análises estatísticas, que incluiu o faseamento computacional dos haplótipos encontrados, avaliação do desequilíbrio de ligação entre os *loci* estudados, verificação das frequências alélicas, aderência ao equilíbrio de Hardy-Weinberg e associação com as características fenotípicas. Para ambos os genes, o programa PHASE (Stephens, Smith et al. 2001) foi utilizado para a reconstrução dos haplótipos. Foi utilizada uma abordagem de faseamento fragmentado para os genes

TYR e *SLC45A2*, em que cada região gênica é inferida separadamente. Para o gene *SLC45A2* as variantes foram agrupadas em seis regiões, uma vez que este gene possui três isoformas descritas (FIGURA 1): uma isoforma com o códon de parada no éxon sete (podendo ser observada uma 3'UTR menor); outra com um códon de parada no éxon seis (3'UTR maior); além da isoforma com um códon de parada no éxon cinco (esta isoforma tem apenas quatro éxons, pois além de não possuir os éxons seis e sete também não possui o éxon três). Dessa forma, as regiões analisadas foram: promotora (comum a todas as isoformas) com 24 pontos de variação; codificante com 7 éxons (isoforma 1) e 32 pontos de variação; codificante com 6 éxons (isoforma 2) e 20 pontos de variação; codificante com 4 éxons (isoforma 3) e 18 pontos de variação, 3'UTR maior (isoforma 2) com 9 pontos de variação e 3'UTR menor (isoforma 1) com dois pontos de variação. Não foi encontrado nenhum ponto de variação na região 5'UTR deste gene e na 3'UTR formada pela isoforma 3 com quatro éxons foi encontrado apenas um ponto de variação. Para o gene *TYR* as variantes foram agrupadas em duas regiões: região promotora com 20 pontos de variação e região codificante com 21 pontos de variação (nenhuma variante foi encontrada nas regiões 5'UTR e 3'UTR). Para ambos os genes (*TYR* e *SLC45A2*) foram considerados os éxons e suas regiões flanqueadoras cobertas por nosso ensaio (íntrons) como região codificante. O teste exato de Fisher bi-caudal, implementado em planilha desenvolvida para esta finalidade (Mendes Junior, C. T.), foi utilizado para averiguar a existência de associações entre alelos, genótipos ou haplótipos e um determinado fenótipo. Para determinar a intensidade desta associação foi realizado o cálculo de *Odds Ratio* (OR) e seu intervalo de confiança (IC) de 95% empregando-se a mesma planilha. O valor OR permite quantificar a magnitude da associação entre um alelo/genótipo/haplótipo e determinado fenótipo por meio da razão entre a probabilidade de um indivíduo portador de um determinado alelo/genótipo/haplótipo apresentar uma determinada característica e a probabilidade de um indivíduo portador do mesmo alelo/genótipo/haplótipo não apresentar tal característica. Foram consideradas significativas as associações com valores de *p* menores ou iguais a 0,05; OR maiores ou iguais a 10 (com algumas exceções que serão discutidas posteriormente) e limite inferior do IC de 95% maior do que um. O desequilíbrio de ligação (DL) foi verificado em ambos os genes (*TYR* e *SLC45A2*) pelo programa Haploview (Barrett, Fry et al. 2005), levando-se em consideração apenas os SNPs. As frequências alélicas, heterozigose observada e esperada, assim como a aderência ao equilíbrio de Hardy-Weinberg foram calculadas pelo programa Arlequin (Excoffier and Lischer 2010) para as 200 amostras coletadas aleatoriamente. Estes cálculos foram realizados *locus-a-locus* para ambos os genes.

RESULTADOS

As características avaliadas na amostra populacional (estratificada em amostra aleatória e amostra direcionada para aumentar os fenótipos de interesse), isto é, cores de olhos, cabelos e pele e presença ou ausência de sardas estão representadas na Tabela 1. Após a análise dos relatórios e arquivos produzidos pelo VCFx e uma verificação ponto a ponto das variantes, foram removidos das análises posteriores os *singletons* (pontos de variação nos quais apenas uma cópia do alelo alternativo estava presente) e as regiões que apresentaram baixa cobertura e *reads* com baixa qualidade. Assim, para o gene *SLC45A2* restaram 58 pontos de variação para as análises posteriores e para o gene *TYR* permaneceram 42 sítios.

Considerando somente os 200 indivíduos coletados aleatoriamente e a definição de *locus* polimórfico como sendo o *locus* que possui o alelo menos frequente com frequência maior que 1%, encontramos 25 variantes (das 58 analisadas) polimórficas para o gene *SLC45A2* e 19 (das 42 analisadas) para o gene *TYR*. Dessa forma, podemos observar nas Tabelas 2 e 3 que o gene *SLC45A2* possui 33 polimorfismos com o alelo menos frequente com frequência entre 0 e 1% (nove deles - rs143386685, rs73077153, rs551953158, rs58784954, rs6451048, rs376784068, rs113928135, rs6874998 e rs45552240 - associados com algum fenótipo avaliado) e o gene *TYR* possui 23 (quatro deles - 88907815, rs142881018, rs55870815, rs75375648 - associados com algum fenótipo avaliado). Em relação ao equilíbrio de Hardy-Weinberg, observou-se que apresentaram desvios as variantes rs35407, rs16891982, rs28777 e rs183671 do gene *SLC45A2* e a variante rs11824466 do gene *TYR*.

O desequilíbrio de ligação foi avaliado para os genótipos não faseados em ambos os genes. Para o gene *SLC45A2* foram considerados 51 SNPs nos 340 indivíduos. Podemos observar na FIGURA 2 dois blocos em desequilíbrio, o primeiro abrangendo parte da região codificante das três isoformas e a 3'UTR das isoformas 2 e 3, composto por 21 pontos de variação e o segundo, abrangendo parte da região promotora, composto por 4 pontos de variação. Já para o gene *TYR* foram considerados 37 SNPs nos 340 indivíduos. Podemos observar na FIGURA 3 dois blocos em desequilíbrio, o primeiro abrangendo parte das regiões promotora e codificante, composto por 17 pontos de variação e o segundo abrangendo parte da região codificante e o SNP 3' *Downstream* composto por 6 pontos de variação.

Com a finalidade de apresentar apenas associações fenotípicas com maior significância estatística, as Tabelas 4 e 5 mostram apenas associações envolvendo 28 variantes (das 58 avaliadas) do gene *SLC45A2* e 15 (das 42 avaliadas) do gene *TYR*,

que respeitem simultaneamente os seguintes critérios: valores de probabilidade obtidos pelo teste exato de Fisher bilateral menores ou iguais a 0,05, valores de *Odds Ratio* (OR) maiores ou iguais a 10 e limite inferior do intervalo de confiança (IC) maiores que um. Com relação às associações encontradas envolvendo haplótipos, divididos de acordo com as regiões gênicas, foram mantidos os mesmos critérios citados acima para selecionar as associações mais fortes (Tabelas 6 e 7).

DISCUSSÃO

A população brasileira é uma das mais heterogêneas em todo o mundo, resultado de cinco séculos de mistura inter-étnica envolvendo principalmente Nativos Americanos, Europeus e Africanos (Salzano and Sans 2014). De acordo com uma estimativa realizada em 2017 baseada no censo de 2010 (IBGE 2010), o Brasil possui mais de 200 milhões de habitantes, dos quais 47,7% se autodeclararam "brancos", 43,1% "pardos", 7,6% "negros" e 1,5% disseram pertencer a outras categorias. Estas proporções variam consideravelmente nas diferentes regiões geográficas brasileiras devido aos aspectos particulares de suas colonizações: houve uma contribuição mais pronunciada dos nativos americanos no Norte, uma alta contribuição africana no Nordeste e uma contribuição relativamente baixa de ambas no Sul (Salzano and Sans 2014). No estado de São Paulo (com aproximadamente 45 milhões de habitantes), as proporções foram de 63,9% autodeclarados "brancos", 29,1% "pardos", 5,5% "negros" e 1,5% de outras categorias. Na cidade de Ribeirão Preto, que conta com 682.302 habitantes, as frequências de "brancos", "pardos" e "negros" relatadas foram de 69,8%, 22,8% e 6,4% respectivamente (IBGE 2010). No presente estudo, dos 204 indivíduos coletados aleatoriamente, a maioria (107 indivíduos) foi classificado com pele moderadamente pigmentada (peles tipo III e IV na escala de Fitzpatrick), seguido por 72 indivíduos de pele clara (peles tipo I e II) e a categoria menos representada foi a de pele escura (peles tipo V e VI, 25 indivíduos). Quando comparamos esses dados com os dados baseados nos critérios adotados pelo IBGE, encontramos: 133 indivíduos classificados pelo grupo de pesquisa como brancos (dentre eles, 70 indivíduos tinham pele clara e 59, pele tipo III); 3 amarelos (dois de pele III); 53 pardos, dentre eles, a grande maioria de pele média (cinco com pele III e 38 com pele IV) e 15 pretos (todos com pele escura). Se levarmos em consideração que a variação na pigmentação da pele é uma característica contínua e que adotar categorias é uma decisão de certa forma arbitrária, a distribuição dos tipos de pele ao longo das categorias baseadas nos critérios do IBGE se mostrou pertinente, havendo um aumento gradual dos tipos de pele mais claros para os mais escuros dos indivíduos

brancos para os pretos. Ainda em relação a amostra coletada aleatoriamente, quatro (rs35407, rs16891982, rs28777 e rs183671) dos 58 *loci* avaliados para o gene *SLC45A2* e um (rs11824466) dos 42 *loci* avaliados para o gene *TYR* apresentaram desvios com relação ao equilíbrio de Hardy-Weinberg. Todos os *loci* que apresentaram desvios possuem excesso de homozigotos em relação ao equilíbrio em ambos os genes. Para o gene *SLC45A2*, era estatisticamente esperado que 2,9 casos apresentassem desvio em 58 análises já que aproximadamente 5% das análises resultam em desvios casuais ao se adotar o nível de significância de 0,05; entretanto, os valores observado e esperado não diferem significativamente ($\chi^2 = 0,4174$; $p = 0,5182$). Contudo, uma vez que estas variantes estão ligadas, um possível fator que pode ter ocasionado esse desvio é a ocorrência de mistura inter-étnica, associada à subestrutura populacional. A alta miscigenação da população de Ribeirão Preto associada à ocorrência de casamentos preferenciais no que se refere a ancestralidade e pigmentação da pele, pode resultar em estratificação populacional (efeito Wallund). Apesar de que evidências de seleção positiva agindo sob o SNP rs16891982, favorecendo fenótipos de pigmentação clara em europeus nos últimos 5000 anos, tenham sido encontradas (Wilde, Timpson et al. 2014), é improvável que sua intensidade seja suficiente para resultar em desvios do equilíbrio de Hardy-Weinberg em uma amostra populacional brasileira. Para o *locus* do gene *TYR*, todavia, é mais provável que esse único desvio seja devido tão somente ao acaso, já que era esperado uma ocorrência de 2,1 casos de desvio em 42 análises.

Para o gene *SLC45A2*, quatro variantes da região promotora foram associadas com fenótipos de pigmentação. Três delas (rs143386685, rs73077153 e rs551953158) possuem alelo alternativo raro e não são consideradas polimórficas na população coletada aleatoriamente (Tabela 2). Já o SNP rs73077154 apresentou seu alelo alternativo associado a fenótipos escuros, assim como todos os outros três alelos alternativos. Os alelos rs73077153*T ($p = 0,0269$, OR = 25,72), rs551953158*T ($p = 0,0269$, OR = 25,72) e rs73077154*C ($p = 2,34 \times 10^{-06}$, OR = 18,78) se mostraram associados com pele escura e o alelo rs143386685*A ($p = 0,0160$, OR = 21,00) com cabelos pretos (Tabela 4). Também foram encontradas 24 variantes associadas com fenótipos nas regiões codificantes e 3'UTR dependendo da isoforma considerada. Treze variantes estão em íntrons (rs250413, rs58784954, rs6451049, rs6451048, rs183671, rs376784068, rs141348470, rs28777, rs1010872, rs113928135, rs35398, rs250416, rs146922924) e uma variante (rs2287949) em éxon nas três isoformas. O SNP rs45552240 encontra-se apenas na região 3'UTR da isoforma1. Dentre as nove variantes restantes, sete (rs6894408, rs6874998, rs55746663, rs57959046, rs35407, rs7380538 e rs35406) estão localizadas no íntron seis da isoforma 1 ou na região

3'UTR da isoforma 2; o SNP rs26722 se localiza no éxon três das isoformas 1 e 2 ou no íntron dois da isoforma 3 e o SNP rs16891982 está no éxon cinco das isoformas 1 e 2 ou na região 3'UTR da isoforma 3.

Em relação as variantes polimórficas localizadas na região intrônica, encontramos fortes associações do alelo rs28777*A com cabelos loiros ($p = 2,83 \times 10^{-24}$, OR = 50,07) e pele clara ($p = 1,01 \times 10^{-26}$, OR = 14,88). Este SNP pertence ao conjunto de SNPs utilizados pelo sistema HirisPlex para predição das cores de olhos e cabelos (Walsh, Liu et al. 2013) e está em forte desequilíbrio de ligação com o SNP rs183671 [LOD (*Log of the Odds*) ≥ 2 e $D' = 0,98$] que mostrou associação do alelo rs183671*T a cabelos pretos ($p = 2,39 \times 10^{-22}$, OR = 29,96) e pele escura ($p = 6,35 \times 10^{-25}$, OR = 12,88) (Tabela 4). O SNP rs26722 é uma mutação não sinônima e causa a substituição do aminoácido lisina para glutamato. No presente estudo seu alelo menos frequente (rs26722*T) se mostrou associado a cabelos pretos ($p = 2,21 \times 10^{-05}$, OR = 18,86), porém, apesar de resultar em substituição de aminoácido, é importante mencionar que este SNP está em forte desequilíbrio de ligação com o SNP rs16891982 [$LOD \geq 2$ e $D' = 0,94$] (FIGURA 2) e esta associação pode não estar relacionada com a determinação da pigmentação e sim estar simplesmente refletindo o desequilíbrio entre eles. Em contrapartida, as associações haplotípicas encontradas em trabalho anterior, oriundo da dissertação de mestrado que inspirou o presente estudo, indicam que a mutação no SNP rs26722 modifica as associações no SNP rs16891982. Foi observado que para as associações significativas com olhos e pele clara que possuíam o alelo rs16891982*G (374Phe), o alelo rs26722*C fortalecia a magnitude das associações (OR maiores) e, similarmente, para as associações do alelo rs16891982*C com fenótipos escuros, a presença do alelo rs26722*T também aumentava a magnitude das associações (Fracasso, de Andrade et al. 2017).

O SNP rs16891982 está localizado no éxon 5 (isoformas 1 e 2) e determina uma mutação não sinônima, causando a substituição do aminoácido leucina para fenilalanina. Este é o SNP mais estudado do gene *SLC45A2*, seus alelos e genótipos já se mostraram associados com a variação normal nas características de pigmentação em vários trabalhos em diferentes populações (Graf, Hodgson et al. 2005, Yuasa, Umetsu et al. 2006, Branicki, Brudnik et al. 2008, Kayser 2015, Fracasso, de Andrade et al. 2017). Tal SNP faz parte do sistema HirisPlex de predição de cores de olhos e cabelos (Walsh, Liu et al. 2013) e já foi associado a risco de desenvolver melanoma (Ibarrola-Villava, Hu et al. 2012, Lopez, Garcia et al. 2014). Em nosso trabalho anterior, seu alelo derivado (rs16891982*G/374Phe) apresentou fortes associações com fenótipos claros: olhos azuis ($p = 1,45 \times 10^{-05}$, OR = 6,74), cabelos loiros ($p = 2,32 \times 10^{-10}$, OR = 11,46) e pele clara ($p = 2,13 \times 10^{-21}$, OR = 5,88) (Fracasso,

de Andrade et al. 2017). Este mesmo alelo foi encontrado em maior frequência na população do atual trabalho (60,75%) e se mostrou fortemente associado (foram as associações mais fortes encontradas) com fenótipos claros: pele clara ($p = 9,54 \times 10^{-35}$, OR = 22,66) e cabelos loiros ($p = 1,59 \times 10^{-26}$, OR = 31,72) (Tabela 4). Este alelo (rs16891982*G/374Phe) é encontrado em alta frequência apenas na população europeia (93,83% - Projeto 1000 Genomas).

Quando observamos os haplótipos inferidos para as regiões codificantes das isoformas 1 e 2 que possuem esse SNP e as associações encontradas, podemos notar que os únicos haplótipos associados com fenótipos de pigmentação claros para ambas isoformas [“Iso2cds1”, pele clara ($p = 3,65 \times 10^{-29}$, OR = 22,63) e cabelos loiros ($p = 6,34 \times 10^{-23}$, OR = 26,19) e “Iso1cds1”, pele clara ($p = 1,08 \times 10^{-24}$, OR = 17,38) e cabelos loiros ($p = 1,94 \times 10^{-22}$, OR = 22,93), Tabela 6] possuem como diferença em relação a todos os outros haplótipos associados com pigmentação escura o alelo rs16891982*G, reforçando as associações encontradas para esse SNP. Os SNPs rs28777, rs183671 e rs26722 do gene *SLC45A2* pertencem ao mesmo bloco e estão em forte desequilíbrio com o SNP rs16891982, porém as fortes associações encontradas para eles indicam que seus alelos podem influenciar a magnitude das associações (FIGURA 2).

Para o gene *TYR*, seis variantes (rs75741664, chr11:88907815, rs142881018, rs61759512, rs11824466 e rs55870815) da região promotora se mostraram associadas com fenótipos de pigmentação. Três delas (chr11:88907815, rs142881018 e rs55870815) foram consideradas monomórficas na população do presente estudo coletada aleatoriamente, sendo que o alelo alternativo (G) da variante chr11:88907815 não havia sido descrito anteriormente, sendo observado apenas na amostra brasileira enriquecida (Tabela 3). Também foram encontradas oito variantes associadas com características de pigmentação na região codificante deste gene, seis localizadas em íntrons (rs11018530, rs67512950, rs12804012, rs75375648, rs113438665 e rs1393350) e duas em éxons (rs1042602 e rs1126809). Também foi encontrado um SNP (rs28521275) localizado na região *downstream* cujo genótipo rs28521275*C/C se mostrou associado com cabelos ruivos ($p = 0,0014$, OR = 10,80) (Tabela 5). Dentre as variantes polimórficas localizadas na região intrônica, o genótipo rs1393350*A/A mostrou associações com olhos azuis ($p = 0,0253$, OR = 13,06) e cabelos castanho-claros ($p = 0,0019$, OR = 16,07), mas seus alelos não apresentaram associações significativas com nenhuma das características avaliadas (Tabela 5). Este SNP faz parte do conjunto de SNPs do sistema HIrisPlex para predição das cores de olhos e cabelos (Walsh, Liu et al. 2013) e já foi associado com habilidade de bronzeamento em euro-americanos em um GWAS (*Genome-Wide Association Study*) (Nan, Kraft et

al. 2009). Neste mesmo trabalho, este SNP se mostrou em forte desequilíbrio de ligação com o SNP exônico rs1126809 ($D' = 1$ e $r^2 = 0,86$). Ambos SNPs também estão em forte desequilíbrio de ligação ($LOD \geq 2$ e $D' = 0,98$) no presente estudo (FIGURA 3). O SNP rs1126809 está localizado no éxon 4 e determina uma mutação não-sinônima, causando a substituição do aminoácido arginina para glutamina. Da mesma forma que para o SNP rs1393350, foram encontradas apenas associações significativas entre seu genótipo rs1126809*A/A e as cores de olhos azuis ($p = 0,0249$, OR = 13,18) e cabelos castanho-claros ($p = 0,0056$, OR = 13,44) (Tabela 5). Esses dados são congruentes com trabalho realizado no Brasil em que o alelo rs1126809*G se mostrou associado com características de pigmentação de pele mais escuras (Cerqueira, Hunemeier et al. 2014). Outro SNP localizado em éxon que se mostrou associado com os fenótipos avaliados foi o rs1042602. Sua mutação também é não-sinônima causando a substituição do aminoácido serina por tirosina. Na presente amostra, apenas o genótipo rs1042602*A/A apresentou associação significativa com a cor de pele clara ($p = 0,0033$, OR = 11,00) (Tabela 5). Ao analisarmos as associações encontradas para os haplótipos inferidos para essa região do gene *TYR*, podemos notar que os haplótipos “cds5” ($p = 1,71 \times 10^{-05}$, OR = 21,2549), “cds7” ($p = 0,0061$, OR = 23,7042) e “cds9” ($p = 0,0017$, OR = 29,2464) estão associados a pele escura e o haplótipo “cds12” ($p = 0,015$, OR = 21,61) a ausência de sardas (Tabela 7). Quando nos atentamos a composição desses haplótipos em relação aos alelos dos SNPs rs1042602, rs1393350 e rs1126809, podemos perceber que todos os haplótipos possuem os alelos referência, reforçando as associações encontradas para eles.

De modo geral, os resultados aqui encontrados reafirmam a importância desempenhada pelos genes *SLC45A2* e *TYR* na geração da diversidade de fenótipos de pigmentação e sugerem sua utilidade na predição fenotípica.

Agradecimentos

O presente trabalho teve apoio financeiro do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, #448242/2014-1) e FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, #2013/15447-0). C.T.M.J. (309572/2014-2), E.A.D. (304931/2014-1), A.L.S. (312547/2009-9) e E.C.C. (304471/2013-5) contam com apoio financeiro colaborativo de CNPq. N.C.A.F. usufruiu de bolsa de doutorado PROEX concedida pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

REFERÊNCIAS BIBLIOGRÁFICAS

- Andrews, S. (2010). "FastQC: a quality control tool for high throughput sequence data." from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bae-Harboe, Y. S. and H. Y. Park (2012). "Tyrosinase: a central regulatory protein for cutaneous pigmentation." *J Invest Dermatol* **132**(12): 2678-2680.
- Barrett, J. C., B. Fry, J. Maller and M. J. Daly (2005). "Haploview: analysis and visualization of LD and haplotype maps." *Bioinformatics* **21**(2): 263-265.
- Bin, B. H., J. Bhin, S. H. Yang, M. Shin, Y. J. Nam, D. H. Choi, D. W. Shin, A. Y. Lee, D. Hwang, E. G. Cho and T. R. Lee (2015). "Membrane-Associated Transporter Protein (MATP) Regulates Melanosomal pH and Influences Tyrosinase Activity." *PLoS One* **10**(6): e0129273.
- Branicki, W., U. Brudnik, J. Draus-Barini, T. Kupiec and A. Wojas-Pelc (2008). "Association of the SLC45A2 gene with physiological human hair colour variation." *J Hum Genet* **53**(11-12): 966-971.
- Castelli, E. C., C. T. Mendes-Junior, A. Sabbagh, I. O. Porto, A. Garcia, J. Ramalho, T. H. Lima, J. D. Massaro, F. C. Dias, C. V. Collares, V. Jamonneau, B. Bucheton, M. Camara and E. A. Donadi (2015). "HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples." *Hum Immunol* **76**(12): 945-953.
- Cerqueira, C. C., T. Hunemeier, J. Gomez-Valdes, V. Ramallo, C. D. Volasko-Krause, A. A. Barbosa, P. Vargas-Pinilla, R. C. Dornelles, D. Longo, F. Rothhammer, G. Bedoya, S. Canizales-Quinteros, V. Acuna-Alonzo, C. Gallo, G. Poletti, R. Gonzalez-Jose, F. M. Salzano, S. M. Callegari-Jacques, L. Schuler-Faccini, A. Ruiz-Linares, M. Catira Bortolini and C. for (2014). "Implications of the admixture process in skin color molecular assessment." *PLoS One* **9**(5): e96886.
- Costin, G. E., J. C. Valencia, W. D. Vieira, M. L. Lamoreux and V. J. Hearing (2003). "Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4." *J Cell Sci* **116**(Pt 15): 3203-3212.
- DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernysky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler and M. J. Daly (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." *Nat Genet* **43**(5): 491-498.
- Dooley, C. M., H. Schwarz, K. P. Mueller, A. Mongera, M. Konantz, S. C. Neuhaus, C. Nusslein-Volhard and R. Geisler (2013). "Slc45a2 and V-ATPase are regulators of melanosomal pH homeostasis in zebrafish, providing a mechanism for human pigment evolution and disease." *Pigment Cell Melanoma Res* **26**(2): 205-217.
- Excoffier, L. and H. E. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." *Mol Ecol Resour* **10**(3): 564-567.
- Ferragina, P., G. Manzini, V. Mäkinen and G. Navarro (2004). An Alphabet-Friendly FM-Index. *String Processing and Information Retrieval*, Springer, Berlin, Heidelberg: 150-160.
- Fitzpatrick, T. B. (1988). "The validity and practicality of sun-reactive skin types I through VI." *Arch Dermatol* **124**(6): 869-871.
- Fracasso, N. C. A., E. S. de Andrade, C. E. V. Wiesel, C. C. F. Andrade, L. R. Zanao, M. S. da Silva, L. A. Marano, E. A. Donadi, C. C. E, A. L. Simoes and C. T. Mendes-

Junior (2017). "Haplotypes from the SLC45A2 gene are associated with the presence of freckles and eye, hair and skin pigmentation in Brazil." *Leg Med (Tokyo)* **25**: 43-51.

Graf, J., R. Hodgson and A. van Daal (2005). "Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation." *Hum Mutat* **25**(3): 278-284.

Hearing, V. J. (2011). "Determination of melanin synthetic pathways." *J Invest Dermatol* **131**(E1): E8-E11.

Ibarrola-Villava, M., H. H. Hu, M. Guedj, L. P. Fernandez, V. Descamps, N. Basset-Seguín, M. Bagot, A. Bensussan, P. Saiag, M. C. Fargnoli, K. Peris, J. A. Aviles, A. Lluch, G. Ribas and N. Soufir (2012). "MC1R, SLC45A2 and TYR genetic variants involved in melanoma susceptibility in southern European populations: results from a meta-analysis." *Eur J Cancer* **48**(14): 2183-2191.

IBGE, I. B. d. G. e. E.-. (2010). 2018, from <http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm>.

Kawakami, A. and D. E. Fisher (2011). "Key discoveries in melanocyte development." *J Invest Dermatol* **131**(E1): E2-4.

Kayser, M. (2015). "Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes." *Forensic Sci Int Genet* **18**: 33-48.

Kondo, T., T. Namiki, S. G. Coelho, J. C. Valencia and V. J. Hearing (2015). "Oculocutaneous albinism: developing novel antibodies targeting the proteins associated with OCA2 and OCA4." *J Dermatol Sci* **77**(1): 21-27.

Lai, X., H. J. Wichers, M. Soler-Lopez and B. W. Dijkstra (2018). "Structure and Function of Human Tyrosinase and Tyrosinase-Related Proteins." *Chemistry* **24**(1): 47-55.

Lassmann, T., Y. Hayashizaki and C. O. Daub (2011). "SAMStat: monitoring biases in next generation sequencing data." *Bioinformatics* **27**(1): 130-131.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." *Bioinformatics* **25**(14): 1754-1760.

Lima, T. H. A., R. V. Buttura, E. A. Donadi, L. C. Veiga-Castelli, C. T. Mendes-Junior and E. C. Castelli (2016). "HLA-F coding and regulatory segments variability determined by massively parallel sequencing procedures in a Brazilian population sample." *Hum Immunol* **77**(10): 841-853.

Liu, F., B. Wen and M. Kayser (2013). "Colorful DNA polymorphisms in humans." *Semin Cell Dev Biol* **24**(6-7): 562-575.

Lopez, S., O. Garcia, I. Yurrebaso, C. Flores, M. Acosta-Herrera, H. Chen, J. Gardeazabal, J. M. Careaga, M. D. Boyano, A. Sanchez, J. A. Raton-Nieto, A. Sevilla, I. Smith-Zubiaga, A. G. de Galdeano, C. Martinez-Cadenas, N. Izagirre, C. de la Rua and S. Alonso (2014). "The interplay between natural selection and susceptibility to melanoma on allele 374F of SLC45A2 gene in a South European population." *PLoS One* **9**(8): e104367.

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." *EMBnet.journal* **17**(1): 10.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." *Genome Res* **20**(9): 1297-1303.

Miller, S. A., D. D. Dykes and H. F. Polesky (1988). "A simple salting out procedure for extracting DNA from human nucleated cells." *Nucleic Acids Res* **16**(3): 1215.

Mondal, M., M. Sengupta and K. Ray (2016). "Functional assessment of tyrosinase variants identified in individuals with albinism is essential for unequivocal determination of genotype-to-phenotype correlation." Br J Dermatol **175**(6): 1232-1242.

Nan, H., P. Kraft, A. A. Qureshi, Q. Guo, C. Chen, S. E. Hankinson, F. B. Hu, G. Thomas, R. N. Hoover, S. Chanock, D. J. Hunter and J. Han (2009). "Genome-wide association study of tanning phenotype in a population of European ancestry." J Invest Dermatol **129**(9): 2250-2257.

Newton, J. M., O. Cohen-Barak, N. Hagiwara, J. M. Gardner, M. T. Davisson, R. A. King and M. H. Brilliant (2001). "Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4." Am J Hum Genet **69**(5): 981-988.

Oetting, W. S. (2000). "The tyrosinase gene and oculocutaneous albinism type 1 (OCA1): A model for understanding the molecular biology of melanin formation." Pigment Cell Res **13**(5): 320-325.

Ramsden, C. A. and P. A. Riley (2014). "Tyrosinase: the four oxidation states of the active site and their relevance to enzymatic activation, oxidation and inactivation." Bioorg Med Chem **22**(8): 2388-2395.

Salzano, F. M. and M. Sans (2014). "Interethnic admixture and the evolution of Latin American populations." Genet Mol Biol **37**(1 Suppl): 151-170.

Solomon, E. I., U. M. Sundaram and T. E. Machonkin (1996). "Multicopper Oxidases and Oxygenases." Chem Rev **96**(7): 2563-2606.

Stephens, M., N. J. Smith and P. Donnelly (2001). "A new statistical method for haplotype reconstruction from population data." Am J Hum Genet **68**(4): 978-989.

Tully, G. (2007). "Genotype versus phenotype: human pigmentation." Forensic Sci Int Genet **1**(2): 105-110.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel and M. A. DePristo (2013). "From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline." Curr Protoc Bioinformatics **43**: 11 10 11-33.

Walsh, S., F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki and M. Kayser (2013). "The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA." Forensic Sci Int Genet **7**(1): 98-115.

Wilde, S., A. Timpson, K. Kirsanow, E. Kaiser, M. Kayser, M. Unterlander, N. Hollfelder, I. D. Potekhina, W. Schier, M. G. Thomas and J. Burger (2014). "Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y." Proc Natl Acad Sci U S A **111**(13): 4832-4837.

Yamaguchi, Y., M. Brenner and V. J. Hearing (2007). "The regulation of skin pigmentation." J Biol Chem **282**(38): 27557-27561.

Yuasa, I., K. Umetsu, S. Harihara, A. Kido, A. Miyoshi, N. Saitou, B. Dashnyam, F. Jin, G. Lucotte, P. K. Chattopadhyay, L. Henke and J. Henke (2006). "Distribution of the F374 allele of the SLC45A2 (MATP) gene and founder-haplotype analysis." Ann Hum Genet **70**(Pt 6): 802-811.

FIGURAS

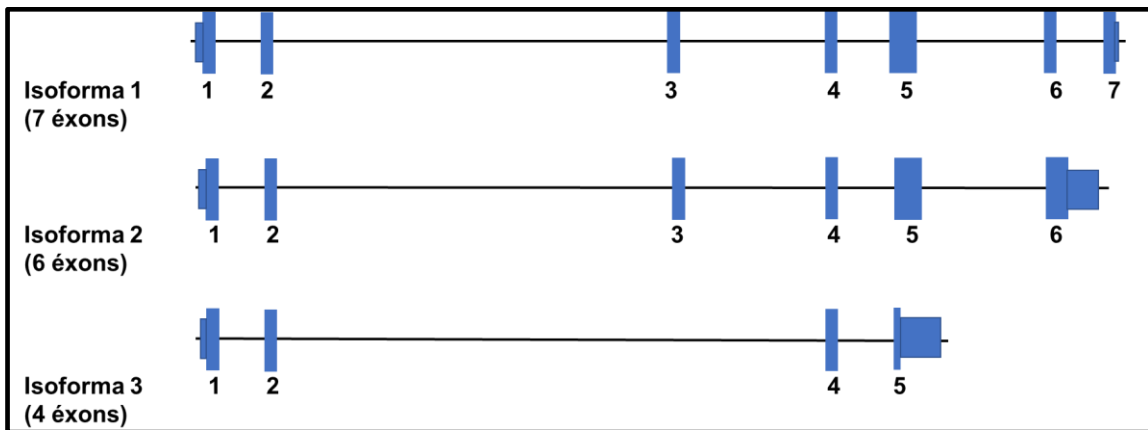


FIGURA 1. Representação esquemática das isoformas do gene *SLC45A2*. Os retângulos azuis representam os éxons e as regiões promotora (à esquerda) e 3'UTR (à direita). A isoforma 1 possui a 3'UTR de menor tamanho (31pb). A isoforma 2 possui 15pb a mais no éxon 6 e a 3'UTR de maior tamanho (1280pb). A isoforma 3 não possui os éxons 3, 6 e 7; o éxon 5 tem apenas 25pb e sua 3'UTR tem 245pb (incluindo parte do éxon 5 das outras duas isoformas). A figura não está em escala.

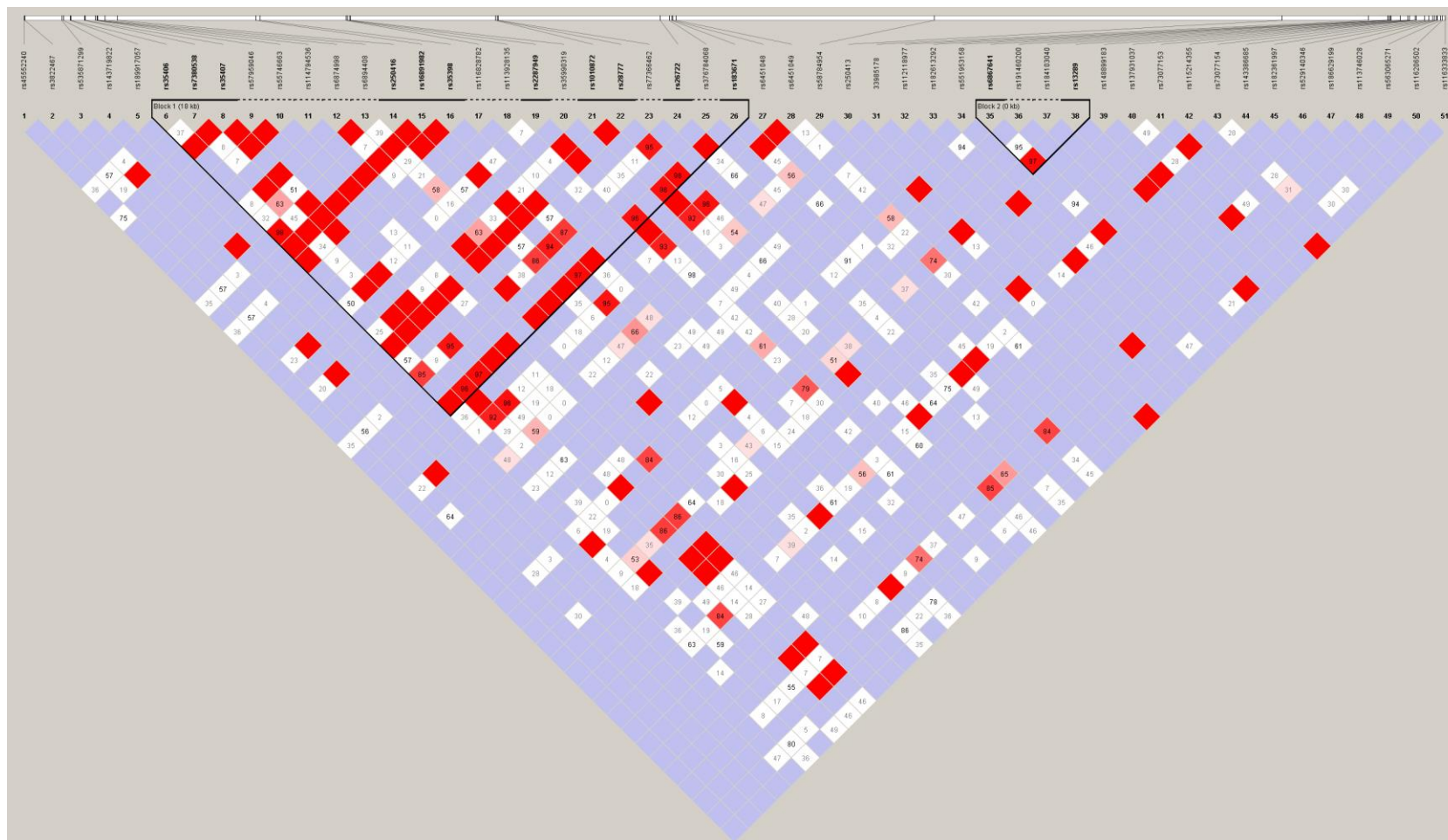


FIGURA 2. Representação do desequilíbrio de ligação para os pontos de variação do gene **SLC45A2**, considerando a região promotora no lado direito da imagem. O bloco em desequilíbrio está delimitado em preto. Vermelho indica LD completo forte [$LOD (Log\ of\ the\ Odds) \geq 2$ e $D' = 1$]; azul indica LD completo fraco [$LOD < 2$ e $D' = 1$]; tons de rosa indicam LD moderado [$LOD \geq 2$ e $D' < 1$]; branco indica ausência de LD [$LOD < 2$ e $D' < 1$]. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). Valores de D' diferentes de 1 estão representados na forma de porcentagem no interior dos quadros com tons de rosa ou brancos.



FIGURA 3. Representação do desequilíbrio de ligação para os pontos de variação do gene *TYR*. O bloco em desequilíbrio está delimitado em preto. Vermelho indica LD completo forte [LOD (Log of the Odds) ≥ 2 e $D' = 1$]; azul indica LD completo fraco [$LOD < 2$ e $D' = 1$]; tons de rosa indicam LD moderado [$LOD \geq 2$ e $D' < 1$]; branco indica ausência de LD [$LOD < 2$ e $D' < 1$]. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). Valores de D' diferentes de 1 estão representados na forma de porcentagem no interior dos quadros com tons de rosa ou brancos.

TABELAS

Tabela 1: Características da amostra populacional de acordo com a heteroclassificação.

Características	Amostras aleatórias		Amostras visando aumentar a amostragem de fenótipos de interesse	
	Mulheres 104	Homens 100	Mulheres 63	Homens 73
Cor dos olhos				
Azul	4	9	12	12
Verde	12	10	18	22
Mel	5	3	6	5
Castanho-claro	28	34	4	7
Castanho-escuro	55	44	23	27
Cor dos cabelos*				
Ruivo	0	0	5	6
Loiro	11	13	17	8
Castanho	79	50	29	28
Preto	13	36	7	30
Cor da pele				
Clara (I + II)	38	34	37	35
Intermediária (III + IV)	53	54	18	15
Escura (V + VI)	13	12	8	23
Sardas				
Presença	84	69	57	47
Ausência	20	31	6	26

*Mulheres que tingem o cabelo e não estavam com a raiz exposta, assim como homens completamente calvos ou com a cabeça raspada ficaram sem classificação nesta categoria, somando um total de oito indivíduos (seis mulheres e dois homens).

Tabela 2: Frequências alélicas, heterozigose observada (H_o), heterozigose esperada (H_e) e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) para cada *locus* do gene *SLC45A2*. Valores significativos ($\alpha = 0,05$) na análise de aderência ao equilíbrio estão marcados em vermelho.

Posição	Região	Locus	Frequência		H_o	H_e	p_{HW}
			Alelo Referência	Alelo Alternativo			
33987471	promotora	rs116333833 (C/A)	0,9950	0,0050	0,0100	0,0100	1,0000
33987394	promotora	rs116206502 (C/A)	0,9700	0,0300	0,0500	0,0584	0,1561
33987346	promotora	rs563065271 (G/A)	0,9950	0,0025	0,0050	0,0050	1,0000
33987237	promotora	rs113746028 (C/T)	0,9750	0,0200	0,0402	0,0395	1,0000
33987219	promotora	rs186629199 (C/G)	0,9925	0,0025	0,0050	0,0050	1,0000
33987154	promotora	rs529140346 (G/T)	0,9950	0,0025	0,0050	0,0050	1,0000
33987110	promotora	rs182361997 (T/A)	0,9925	0,0050	0,0101	0,0100	1,0000
33986992	promotora	rs143386685 (G/A)	0,9950	0,0050	0,0100	0,0100	1,0000
33986871	promotora	rs73077154 (G/C)	0,9725	0,0275	0,0550	0,0536	1,0000
33986606	promotora	rs115214355 (T/C)	0,9950	0,0050	0,0100	0,0100	1,0000
33986605	promotora	rs73077153 (C/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33986567	promotora	rs137931037 (C/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33986410	promotora	rs148899183 (G/A)	0,9975	0,0025	0,0050	0,0050	1,0000
33986409	promotora	rs13289 (C/G)	0,5275	0,4725	0,4350	0,4997	0,0874
33986343	promotora	rs184103040 (G/A)	0,9925	0,0050	0,0101	0,0100	1,0000
33986123	promotora	rs191460200 (G/T)	0,9925	0,0075	0,0150	0,0149	1,0000
33985861	promotora	rs67752819 (CATT/C)	0,5325	0,4650	0,4472	0,4988	0,1585
33985857	promotora	rs6867641 (T/C)	0,3425	0,6550	0,4271	0,4526	0,4371
33985820	promotora	rs551953158 (C/T)	1,0000	0,0000	-	-	-
33985792	promotora	rs182613292 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
33985753	promotora	rs112118977 (T/A)	0,9900	0,0050	0,0101	0,0101	1,0000
33985726	promotora	rs369350969 (TTC/T)	0,9950	0,0050	0,0100	0,0100	1,0000
33985178	promotora	33985178 (A/G)	1,0000	0,0000	-	-	-
33984981	promotora	rs548644095 (TTCTC/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33982568	intron 1	rs250413 (G/A)	0,9625	0,0350	0,0603	0,0681	0,2118
33972116	intron 2	rs58784954 (G/A)	0,9925	0,0075	0,0150	0,0149	1,0000
33964352	intron 2	rs6451049 (T/C)	0,0650	0,9250	0,1320	0,1236	1,0000
33964241	intron 2	rs6451048 (A/G)	0,0025	0,9925	0,0051	0,0051	1,0000
33964210	intron 2	rs183671 (T/G)	0,3325	0,6625	0,3737	0,4456	0,0261

33964140	intron 2	rs376784068 (T/A)	0,9925	0,0025	0,0051	0,0051	1,0000
33963870	exon 3 (iso1 e 2)/intron 2 (iso3)	rs26722 (C/T)	0,8800	0,1200	0,2200	0,2117	0,7460
33963731	intron 3 (iso1 e 2)/intron 2 (iso3)	rs141348470 (AAAC/A)	0,0700	0,9275	0,1407	0,1311	0,6047
33958993	intron 3 (iso1 e 2)/intron 2 (iso3)	rs77366462 (G/A)	0,9950	0,0025	0,0050	0,0050	1,0000
33958959	intron 3 (iso1 e 2)/intron 2 (iso3)	rs28777 (C/A)	0,3250	0,6725	0,3568	0,4392	0,0103
33958910	intron 3 (iso1 e 2)/intron 2 (iso3)	rs1010872 (A/C)	0,0750	0,9250	0,1500	0,1391	0,6071
33958873	intron 3 (iso1 e 2)/intron 2 (iso3)	rs147962674 (GTTGTATTTATTT/G)	0,9900	0,0100	0,0200	0,0199	1,0000
33954546	exon 4	rs35990319 (G/A)	1,0000	0,0000	-	-	-
33954511	exon 4	rs2287949 (T/C)	0,0650	0,9350	0,1300	0,1219	1,0000
33954457	intron 4	rs113928135 (G/A)	0,9900	0,0100	0,0200	0,0199	1,0000
33954408	intron 4	rs116828782 (G/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33954387	intron 4	rs560973795 (TAGCCCA/T)	0,9950	0,0050	0,0100	0,0100	1,0000
33951826	intron 4	rs35398 (G/T)	0,8950	0,1050	0,1900	0,1884	1,0000
33951693	exon 5 (iso1 e 2)/3'UTR (iso3)	rs16891982 (C/G)	0,3925	0,6075	0,4050	0,4781	0,0374
33947544	intron 5 (iso1 e 2)	rs250416 (C/A)	0,8850	0,1100	0,2020	0,1980	1,0000
33947168	intron 6 (iso1)/3'UTR (iso2)	rs6894408 (C/G)	0,9775	0,0225	0,0450	0,0441	1,0000
33946949	intron 6 (iso1)/3'UTR (iso2)	rs6874998 (T/C)	0,9900	0,0100	0,0200	0,0199	1,0000
33946946	intron 6 (iso1)/3'UTR (iso2)	rs114794536 (T/A)	0,9800	0,0200	0,0400	0,0393	1,0000
33946931	intron 6 (iso1)/3'UTR (iso2)	rs55746663 (A/G)	0,9675	0,0325	0,0650	0,0631	1,0000
33946876	intron 6 (iso1)/3'UTR (iso2)	rs57959046 (C/A)	0,9675	0,0325	0,0650	0,0631	1,0000
33946571	intron 6 (iso1)/3'UTR (iso2)	rs35407 (A/G)	0,2975	0,7025	0,3350	0,4190	0,0068
33946555	intron 6 (iso1)/3'UTR (iso2)	rs7380538 (A/G)	0,0825	0,9175	0,1550	0,1518	1,0000
33946143	intron 6 (iso1)/3'UTR (iso2)	rs35406 (A/G)	0,8950	0,1050	0,1900	0,1884	1,0000
33946117	intron 6 (iso1)/3'UTR (iso2)	rs189917057 (A/T)	0,9950	0,0050	0,0100	0,0100	1,0000
33945915	intron6 (iso1)	rs143719822 (T/A)	0,9925	0,0050	0,0101	0,0100	1,0000
33945902	intron6 (iso1)	rs146922924 (C/CA)	0,9700	0,0275	0,0553	0,0539	1,0000
33945869	intron6 (iso1)	rs535871299 (C/T)	0,9900	0,0050	0,0101	0,0101	1,0000
33944752	3'UTR (iso1)	rs3822467 (C/T)	0,9975	0,0025	0,0050	0,0050	1,0000
33944729	3'UTR (iso1)	rs45552240 (T/G)	0,9975	0,0025	0,0050	0,0050	1,0000

Tabela 3: Frequências alélicas, heterozigose observada (H_o), heterozigose esperada (H_e) e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) para cada *locus* do gene *TYR*. Valores significativos ($\alpha = 0,05$) na análise de aderência ao equilíbrio estão marcados em vermelho.

Posição	Região	Locus	Frequência		H_o	H_e	p_{HW}
			Alelo Referência	Alelo Alternativo			
88907621	promotora	rs7924925 (G/A)	0,1525	0,8475	0,2650	0,2591	1,0000
88907737	promotora	rs151276292 (A/C)	0,9950	0,0050	0,0100	0,0100	1,0000
88907781	promotora	rs75741664 (C/T)	0,9875	0,0125	0,0250	0,0248	1,0000
88907811	promotora	rs181833244 (A/G)	0,9950	0,0050	0,0100	0,0100	1,0000
88907996	promotora	rs7925404 (G/A)	0,7350	0,2650	0,3800	0,3905	0,7166
88908275	promotora	rs142881018 (CCTGA/C)	0,9900	0,0100	0,0200	0,0199	1,0000
88908457	promotora	rs139927890 (A/AAT)	0,9875	0,0125	0,0250	0,0248	1,0000
88908667	promotora	rs61759511 (T/C)	0,9900	0,0100	0,0200	0,0199	1,0000
88908789	promotora	rs61759512 (G/A)	0,9825	0,0125	0,0253	0,0250	1,0000
88908905	promotora	rs11824466 (C/T)	0,9800	0,0200	0,0200	0,0393	0,0014
88909024	promotora	rs114470821 (A/G)	0,9975	0,0025	0,0050	0,0050	1,0000
88909316	promotora	rs61759516 (G/A)	0,9875	0,0100	0,0201	0,0200	1,0000
88909753	promotora	rs1160550168 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
88909835	promotora	rs563971364 (TA/T)	0,9975	0,0025	0,0050	0,0050	1,0000
88910428	promotora	rs143441124 (C/A)	0,9950	0,0025	0,0050	0,0050	1,0000
88910589	promotora	rs5021654 (G/C)	0,5800	0,4175	0,4472	0,4883	0,2454
88910821	promotora	rs4547091 (C/T)	0,5800	0,4175	0,4472	0,4883	0,2466
88910882	promotora	rs55870815 (T/C)	0,9950	0,0025	0,0050	0,0050	1,0000
88910923	promotora	rs1799989 (C/A)	0,8750	0,1250	0,2200	0,2193	1,0000
88911235	exon 1	rs1939261 (G/A)	0,9925	0,0075	0,0150	0,0149	1,0000
88911299	exon 1	rs1939260 (C/T)	0,9925	0,0075	0,0150	0,0149	1,0000
88911696	exon 1	rs1042602 (C/A)	0,6750	0,3250	0,4100	0,4399	0,3395
88924385	exon 2	rs4987234 (T/C)	0,9925	0,0075	0,0150	0,0149	1,0000
88924611	intron 2	rs199569456 (A/AT)	0,9975	0,0025	0,0050	0,0050	1,0000
88924613	intron 2	rs142575793 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
88933411	intron 2	rs138538705 (T/G)	0,9400	0,0200	0,0430	0,0422	1,0000
88933629	intron 2	rs11018530 (C/G)	0,8450	0,1525	0,2663	0,2602	1,0000
88933658	intron 2	rs141481082 (T/C)	0,9900	0,0075	0,0151	0,0150	1,0000
88933717	intron 2	rs67512950 (ACTT/A)	0,8600	0,1400	0,2500	0,2414	0,7733
88960790	intron 2	rs12804012 (G/A)	0,8325	0,1650	0,2915	0,2774	0,6104

88960816	intron 2	rs75375648 (T/C)	0,9900	0,0075	0,0151	0,0150	1,0000
88960914	intron 2	rs113438665 (T/C)	0,9850	0,0125	0,0251	0,0249	1,0000
88961165	intron 3	rs147210895 (C/T)	0,9800	0,0050	0,0103	0,0103	1,0000
88961188	intron 3	rs3793975 (G/A)	0,8600	0,1250	0,2268	0,2251	1,0000
89010998	intron 3	rs184834824 (C/T)	0,9950	0,0050	0,0100	0,0100	1,0000
89011046	intron 3	rs1393350 (G/A)	0,8200	0,1800	0,2700	0,2959	0,2320
89017961	exon 4	rs1126809 (G/A)	0,8025	0,1850	0,2806	0,3039	0,3433
89018281	intron 4	rs145841599 (G/A)	0,9575	0,0075	0,0160	0,0160	1,0000
89028204	intron 4	rs561489771 (C/G)	0,9950	0,0050	0,0100	0,0100	1,0000
89028323	exon 5	rs13312744 (T/C)	0,9975	0,0025	0,0050	0,0050	1,0000
89028969	3' <i>Downstream</i>	rs28521275 (C/T)	0,4475	0,4775	0,5202	0,5010	0,6510

Tabela 4: Alelos e genótipos de SNPs/Indels do gene *SLC45A2* associados com as características fenotípicas avaliadas.

SNPs/Indels	Alelo/ Genótipo	Característica	Valor de <i>p</i>	Odds Ratio (IC 95%)
rs4552240	T/G	cabelos cast claro x todos	0,0193	14,3793 (1,4696 - 140,6928)
	G	cabelos cast claro x todos	0,0197	14,0420 (1,4481 - 136,1646)
rs146922924	C/C	pele clara x pele escura	0,0001	27,1915 (3,3560 - 220,3131)
	C	pele clara x pele escura	0,0001	24,9029 (3,1166 - 198,9831)
	C/C	pele clara x pele média	0,0048	11,0938 (1,4006 - 87,8686)
	C	pele clara x pele média	0,0052	10,7143 (1,3623 - 84,2666)
	C/C	pele clara x todos	0,0003	15,4171 (2,0390 - 116,5719)
	C	pele clara x todos	0,0003	14,6748 (1,9529 - 110,2713)
	C/CA	pele escura x pele clara	0,0001	27,1915 (3,3560 - 220,3131)
	CA	pele escura x pele clara	0,0001	24,9029 (3,1166 - 198,9831)
	C/CA	pele média x pele clara	0,0048	11,0938 (1,4006 - 87,8686)
	CA	pele média x pele clara	0,0052	10,7143 (1,3623 - 84,2666)
rs35406	A/G	cabelos cast escuro x loiros	0,0005	14,5000 (1,9171 - 109,6731)
	G	cabelos cast escuro x loiros	0,0001	15,7907 (2,1324 - 116,9318)
	A/G	cabelos pretos x loiros	0,0002	16,5000 (2,1484 - 126,7248)
	G	cabelos pretos x loiros	0,0001	15,7297 (2,0935 - 118,1851)
	A/G	cabelos cast claro x loiros	0,0111	10,5600 (1,3126 - 84,9573)
	G	cabelos cast claro x loiros	0,0039	11,5688 (1,4859 - 90,0693)
	A/A	cabelos loiros x cast escuro	0,0002	16,5161 (2,1896 - 124,5792)
	A	cabelos loiros x cast escuro	0,0001	15,7907 (2,1324 - 116,9318)
	A/A	cabelos loiros x pretos	0,0001	17,5238 (2,2854 - 134,3689)
	A	cabelos loiros x pretos	0,0001	15,7297 (2,0935 - 118,1851)
	A/A	cabelos loiros x cast claro	0,0057	11,7551 (1,4708 - 93,9488)
	A	cabelos loiros x cast claro	0,0039	11,5688 (1,4859 - 90,0693)
	A/A	cabelos loiros x todos	0,0002	14,6368 (1,9832 - 108,0224)
	A	cabelos loiros x todos	0,0001	13,9116 (1,9107 - 101,2888)
rs7380538	G/G	pele clara x pele escura	6,55 x 10⁻¹²	19,9333 (7,5448 - 52,6637)
	G	pele clara x pele escura	6,63 x 10⁻¹²	15,6667 (6,2762 - 39,1069)
	A/G	pele escura x pele clara	1,25 x 10⁻¹⁰	17,2500 (6,5147 - 45,6754)
	A	pele escura x pele clara	6,63 x 10⁻¹²	15,6667 (6,2762 - 39,1069)
	A/A	pele escura x todos	0,0267	26,1009 (1,2359 - 551,2192)
	A/G	cabelos pretos x loiros	6,85 x 10⁻⁰⁶	24,4211 (3,2070 - 185,9645)
	A	cabelos pretos x loiros	1,85 x 10⁻⁰⁶	23,0288 (3,0971 - 171,2306)
	G/G	cabelos loiros x pretos	1,50 x 10⁻⁰⁶	27,0545 (3,5582 - 205,7092)
	G	cabelos loiros x pretos	1,85 x 10⁻⁰⁶	23,0288 (3,0971 - 171,2306)
	G/G	cabelos loiros x todos	0,0014	11,4383 (1,5455 - 84,6542)
G	cabelos loiros x todos	0,0015	10,7366 (1,4697 - 78,4362)	
rs35407	G/G	pele clara x pele escura	7,10 x 10⁻¹⁸	23,7931 (10,1482 - 55,7847)
	G	pele clara x pele escura	6,25 x 10⁻²⁵	13,4471 (7,9405 - 22,7722)
	A/A	pele escura x pele clara	3,05 x 10⁻¹¹	20,8500 (7,3893 - 58,8313)
	A	pele escura x pele clara	6,25 x 10⁻²⁵	13,4471 (7,9405 - 22,7722)
	A/A	olhos cast escuro x verdes	2,25 x 10⁻⁰⁵	11,0092 (2,5704 - 47,1525)
	A/A	olhos cast escuro x azuis	0,0007	13,2110 (1,7529 - 99,5653)
	A/G	cabelos cast escuro x loiros	2,83 x 10⁻⁰⁶	10,5676 (3,1162 - 35,8359)
	A	cabelos cast escuro x loiros	4,00 x 10⁻¹⁰	14,6296 (4,4961 - 47,6025)
	A/A	cabelos pretos x loiros	1,71 x 10⁻⁰⁷	53,4425 (3,1843 - 896,9244)
	A/G	cabelos pretos x loiros	1,53 x 10⁻⁰⁶	12,1389 (3,5022 - 42,0745)
	A	cabelos pretos x loiros	7,16 x 10⁻²²	41,9369 (12,7799 - 137,6146)
	A	cabelos pretos x ruivos	1,66 x 10⁻⁰⁵	13,2432 (3,0010 - 58,4426)
	G/G	cabelos loiros x cast escuro	3,33 x 10⁻⁰⁹	16,6111 (4,9067 - 56,2354)
	G	cabelos loiros x cast escuro	4,00 x 10⁻¹⁰	14,6296 (4,4961 - 47,6025)
	G/G	cabelos loiros x pretos	2,08 x 10⁻¹⁷	57,9259 (16,1339 - 207,9728)
G	cabelos loiros x pretos	7,16 x 10⁻²²	41,9369 (12,7799 - 137,6146)	

	G/G	cabelos loiros x todos	1,79 x 10⁻¹¹	18,4697 (5,6161 - 60,7409)
	G	cabelos loiros x todos	1,36 x 10⁻¹³	17,8763 (5,5945 - 57,1212)
	G/G	cabelos ruivos x pretos	1,20 x 10⁻⁰⁴	17,0000 (3,3715 - 85,7187)
	G	cabelos ruivos x pretos	1,66 x 10⁻⁰⁵	13,2432 (3,0010 - 58,4426)
rs57959046	C/C	pele clara x pele escura	3,60 x 10⁻⁰⁶	34,9556 (4,3918 - 278,2231)
	C	pele clara x pele escura	1,48 x 10⁻⁰⁶	34,4400 (4,4219 - 268,2369)
	C/C	pele clara x pele média	0,0013	13,4063 (1,7192 - 104,5441)
	C	pele clara x pele média	0,0014	12,8507 (1,6597 - 99,5022)
	C/C	pele clara x todos	2,31 x 10⁻⁰⁵	19,0116 (2,5363 - 142,5058)
	C	pele clara x todos	1,74 x 10⁻⁰⁵	18,7174 (2,5171 - 139,1815)
	C/A	pele escura x pele clara	1,36 x 10⁻⁰⁵	31,087 (3,8748 - 249,4078)
	A	pele escura x pele clara	1,48 x 10⁻⁰⁶	34,4400 (4,4219 - 268,2369)
	C/A	pele média x pele clara	0,0013	13,4063 (1,7192 - 104,5441)
A	pele média x pele clara	0,0014	12,8507 (1,6597 - 99,5022)	
rs55746663	A/A	pele clara x pele escura	9,32 x 10⁻⁰⁷	39,0000 (4,9321 - 308,3854)
	A	pele clara x pele escura	4,10 x 10⁻⁰⁷	37,6869 (4,8675 - 291,795)
	A/A	pele clara x pele média	0,0013	13,4063 (1,7192 - 104,5441)
	A	pele clara x pele média	0,0014	12,8507 (1,6597 - 99,5022)
	A/A	pele clara x todos	2,33 x 10⁻⁰⁵	19,9535 (2,6665 - 149,3128)
	A	pele clara x todos	1,71 x 10⁻⁰⁵	19,5504 (2,6334 - 145,1449)
	A/G	pele escura x pele clara	3,60 x 10⁻⁰⁶	34,9556 (4,3918 - 278,2231)
	G	pele escura x pele clara	4,10 x 10⁻⁰⁷	37,6869 (4,8675 - 291,795)
	A/G	pele média x pele clara	0,0013	13,4063 (1,7192 - 104,5441)
G	pele média x pele clara	0,0014	12,8507 (1,6597 - 99,5022)	
rs6874998	T/T	pele clara x pele escura	0,0057	24,7714 (1,3112 - 467,9935)
	T	pele clara x pele escura	0,0059	23,9309 (1,2778 - 448,1989)
	T/C	pele escura x pele clara	0,0057	24,7714 (1,3112 - 467,9935)
	C	pele escura x pele clara	0,0059	23,9309 (1,2778 - 448,1989)
rs6894408	C/C	pele clara x pele escura	2,58 x 10⁻⁰⁵	50,6495 (2,8699 - 893,8886)
	C	pele clara x pele escura	3,14 x 10⁻⁰⁵	46,9330 (2,6853 - 820,2859)
	C/C	pele clara x pele média	0,0031	18,5396 (1,0597 - 324,3435)
	C	pele clara x pele média	0,0033	17,9982 (1,0339 - 313,3240)
	C/C	pele clara x todos	0,0001	26,4183 (1,5716 - 444,0963)
	C	pele clara x todos	0,0002	25,2869 (1,5108 - 423,2455)
	C/G	pele escura x pele clara	2,58 x 10⁻⁰⁵	50,6495 (2,8699 - 893,8886)
	G	pele escura x pele clara	3,14 x 10⁻⁰⁵	46,9330 (2,6853 - 820,2859)
	C/G	pele média x pele clara	0,0031	18,5396 (1,0597 - 324,3435)
G	pele média x pele clara	0,0033	17,9982 (1,0339 - 313,324)	
rs250416	C/A	cabelos cast escuro x loiros	0,0002	16,1720 (2,1431 - 122,0352)
	A	cabelos cast escuro x loiros	3,68 x 10⁻⁰⁵	17,0283 (2,3039 - 125,8579)
	C/A	cabelos pretos x loiros	0,0002	16,1563 (2,1027 - 124,1365)
	A	cabelos pretos x loiros	0,0003	15,4054 (2,0499 - 115,7723)
	C/A	cabelos cast claro x loiros	0,0057	11,7500 (1,4690 - 93,9834)
	A	cabelos cast claro x loiros	0,0073	10,5556 (1,3473 - 82,6977)
	C/C	cabelos loiros x cast escuro	4,35 x 10⁻⁰⁵	18,2778 (2,4275 - 137,6212)
	C	cabelos loiros x cast escuro	3,68 x 10⁻⁰⁵	17,0283 (2,3039 - 125,8579)
	C/C	cabelos loiros x pretos	0,0001	17,1587 (2,2368 - 131,6245)
	C	cabelos loiros x pretos	0,0003	15,4054 (2,0499 - 115,7723)
	C/C	cabelos loiros x cast claro	0,0057	11,7500 (1,4690 - 93,9834)
	C	cabelos loiros x cast claro	0,0073	10,5556 (1,3473 - 82,6977)
	C/C	cabelos loiros x todos	0,0001	15,2374 (2,0648 - 112,4451)
C	cabelos loiros x todos	0,0001	14,1089 (1,9380 - 102,7123)	
rs16891982	G/G	pele clara x pele escura	6,27 x 10⁻²⁴	171,2857 (22,8580 - 1283,5172)
	G	pele clara x pele escura	9,54 x 10⁻³⁵	22,6646 (12,9041 - 39,8078)
	G/G	pele clara x pele média	7,08 x 10⁻¹⁹	10,0978 (5,8536 - 17,4193)
	G/G	pele clara x todos	5,42 x 10⁻²⁸	14,8387 (8,7276 - 25,2288)
	C/C	pele escura x pele clara	6,08 x 10⁻¹⁷	30,2468 (11,9371 - 76,6407)
	C	pele escura x pele clara	9,54 x 10⁻³⁵	22,6646 (12,9041 - 39,8078)

	G/G	pele média x pele escura	0,0001	16,9626 (2,2596 - 127,3389)
	C/C	olhos cast escuro x azuis	6,12 x 10⁻⁰⁶	21,6774 (2,8916 - 162,5095)
	C/G	cabelos cast escuro x loiros	1,20 x 10⁻⁰⁷	13,2736 (3,9201 - 44,9449)
	C/C	cabelos pretos x cast claro	8,85 x 10⁻⁰⁸	10,6909 (3,9024 - 29,2884)
	C/C	cabelos pretos x loiros	1,21 x 10⁻⁰⁸	22,4318 (5,1222 - 98,2367)
	C/G	cabelos pretos x loiros	1,53 x 10⁻⁰⁶	12,1389 (3,5022 - 42,0745)
	C	cabelos pretos x loiros	1,59 x 10⁻²⁶	31,7200 (13,7453 - 73,2000)
	C/C	cabelos pretos x ruivos	0,0021	21,9663 (1,2549 - 384,5092)
	C	cabelos pretos x ruivos	2,82 x 10⁻⁰⁷	15,4533 (4,3775 - 54,5527)
	G/G	cabelos cast claro x pretos	1,29 x 10⁻¹²	20,5556 (7,7474 - 54,5387)
	G/G	cabelos loiros x cast escuro	2,21 x 10⁻¹¹	16,2000 (5,9886 - 43,8232)
	G/G	cabelos loiros x pretos	2,88 x 10⁻²³	117,3333 (33,8689 - 406,4822)
	G	cabelos loiros x pretos	1,59 x 10⁻²⁶	31,7200 (13,7453 - 73,2000)
	G/G	cabelos loiros x todos	8,09 x 10⁻¹⁴	17,0667 (6,5593 - 44,4060)
	G	cabelos loiros x todos	7,07 x 10⁻¹⁵	10,7925 (4,9183 - 23,6825)
	G/G	cabelos ruivos x pretos	2,84 x 10⁻⁰⁶	35,5556 (7,4349 - 170,0346)
	G	cabelos ruivos x pretos	2,82 x 10⁻⁰⁷	15,4533 (4,3775 - 54,5527)
rs35398	G/T	cabelos cast escuro x loiros	0,0005	14,5000 (1,9171 - 109,6731)
	T	cabelos cast escuro x loiros	0,0001	15,7907 (2,1324 - 116,9318)
	G/T	cabelos pretos x loiros	0,0002	16,5000 (2,1484 - 126,7248)
	T	cabelos pretos x loiros	0,0001	15,7297 (2,0935 - 118,1851)
	G/T	cabelos cast claro x loiros	0,0057	11,7551 (1,4708 - 93,9488)
	T	cabelos cast claro x loiros	0,0073	10,5818 (1,3512 - 82,8737)
	G/G	cabelos loiros x cast escuro	0,0002	16,5161 (2,1896 - 124,5792)
	G	cabelos loiros x cast escuro	0,0001	15,7907 (2,1324 - 116,9318)
	G/G	cabelos loiros x pretos	0,0001	17,5238 (2,2854 - 134,3689)
	G	cabelos loiros x pretos	0,0001	15,7297 (2,0935 - 118,1851)
	G/G	cabelos loiros x cast claro	0,0057	11,7551 (1,4708 - 93,9488)
	G	cabelos loiros x cast claro	0,0073	10,5818 (1,3512 - 82,8737)
	G/G	cabelos loiros x todos	0,0002	14,6368 (1,9832 - 108,0224)
G	cabelos loiros x todos	0,0001	13,6941 (1,8805 - 99,7234)	
rs113928135	G/A	cabelos pretos x todos	0,0156	12,3415 (1,3601 - 111,9866)
	A	cabelos pretos x todos	0,0160	12,0714 (1,3399 - 108,7534)
rs2287949	C/C	pele clara x pele escura	8,89 x 10⁻⁰⁹	15,4444 (5,4247 - 43,9711)
	C	pele clara x pele escura	3,33 x 10⁻⁰⁸	12,3043 (4,4913 - 33,7091)
	T/C	pele escura x pele clara	8,89 x 10⁻⁰⁹	15,4444 (5,4247 - 43,9711)
	T	pele escura x pele clara	3,33 x 10⁻⁰⁸	12,3043 (4,4913 - 33,7091)
	T/C	cabelos pretos x loiros	2,91 x 10⁻⁰⁵	20,8000 (2,7235 - 158,8556)
	T	cabelos pretos x loiros	0,0001	17,2740 (2,3059 - 129,4033)
	C/C	cabelos loiros x pretos	2,91 x 10⁻⁰⁵	20,8000 (2,7235 - 158,8556)
C	cabelos loiros x pretos	0,0001	17,2740 (2,3059 - 129,4033)	
rs1010872	C/C	pele clara x pele escura	6,83 x 10⁻¹²	22,4194 (7,9552 - 63,1824)
	C	pele clara x pele escura	2,01 x 10⁻¹¹	17,1116 (6,3770 - 45,9161)
	A/C	pele escura x pele clara	3,05 x 10⁻¹¹	20,8500 (7,3893 - 58,8313)
	A	pele escura x pele clara	2,01 x 10⁻¹¹	17,1116 (6,3770 - 45,9161)
	A/C	cabelos pretos x loiros	6,85 x 10⁻⁰⁶	24,4211 (3,2070 - 185,9645)
	A	cabelos pretos x loiros	6,33 x 10⁻⁰⁶	21,3262 (2,8631 - 158,8528)
	C/C	cabelos loiros x pretos	3,15 x 10⁻⁰⁶	25,7143 (3,3795 - 195,6575)
	C	cabelos loiros x pretos	6,33 x 10⁻⁰⁶	21,3262 (2,8631 - 158,8528)
	C/C	cabelos loiros x todos	0,0015	10,9367 (1,4769 - 80,9906)
	C	cabelos loiros x todos	0,0023	10,1233 (1,3845 - 74,0222)
rs28777	A/A	pele clara x pele escura	4,91 x 10⁻¹⁹	30,7611 (12,0348 - 78,6257)
	A	pele clara x pele escura	1,01 x 10⁻²⁶	14,8810 (8,7413 - 25,3329)
	C/C	pele escura x pele clara	4,77 x 10⁻¹²	20,4713 (7,73 - 54,2135)
	C	pele escura x pele clara	1,01 x 10⁻²⁶	14,8810 (8,7413 - 25,3329)
	C/C	olhos cast escuro x azuis	0,0004	14,8077 (1,9667 - 111,4887)
	C/A	cabelos cast escuro x loiros	3,05 x 10⁻⁰⁶	10,3378 (3,0462 - 35,0835)
	C	cabelos cast escuro x loiros	5,21 x 10⁻¹¹	15,9697 (4,9116 - 51,9244)

	C/C	cabelos pretos x loiros	2,98 x 10⁻⁰⁸	61,8952 (3,6909 - 1037,9587)
	C/A	cabelos pretos x loiros	7,20 x 10⁻⁰⁷	12,7174 (3,665 - 44,1283)
	C	cabelos pretos x loiros	2,83 x 10⁻²⁴	50,0769 (15,2257 - 164,7022)
	C	cabelos pretos x ruivos	2,83 x 10⁻⁰⁶	16,1538 (3,6549 - 71,3970)
	A/A	cabelos cast claro x pretos	4,45 x 10⁻¹⁰	10,5495 (4,7766 - 23,2992)
	A/A	cabelos loiros x cast escuro	1,03 x 10⁻⁰⁹	17,8947 (5,2802 - 60,6461)
	A	cabelos loiros x cast escuro	5,21 x 10⁻¹¹	15,9697 (4,9116 - 51,9244)
	A/A	cabelos loiros x pretos	5,07 x 10⁻²⁰	83,0769 (22,4278 - 307,7333)
	A	cabelos loiros x pretos	2,83 x 10⁻²⁴	50,0769 (15,2257 - 164,7022)
	A/A	cabelos loiros x todos	6,65 x 10⁻¹²	20,0806 (6,0990 - 66,1141)
	A	cabelos loiros x todos	1,18 x 10⁻¹⁴	19,3641 (6,0595 - 61,8816)
	A/A	cabelos ruivos x pretos	1,58 x 10⁻⁰⁵	24,9231 (4,8246 - 128,7495)
	A	cabelos ruivos x pretos	2,83 x 10⁻⁰⁶	16,1538 (3,6549 - 71,3970)
rs141348470	A/A	pele clara x pele escura	8,57 x 10⁻⁰⁸	10,1250 (4,1365 - 24,7834)
	AAAC/A	pele escura x pele clara	8,57 x 10⁻⁰⁸	10,1250 (4,1365 - 24,7834)
	AAAC/A	cabelos pretos x loiros	0,0001	10,5254 (2,3789 - 46,5696)
	A/A	cabelos loiros x pretos	0,0001	10,5254 (2,3789 - 46,5696)
rs26722	C/T	cabelos cast escuro x loiros	0,0002	15,8298 (2,0969 - 119,5036)
	T	cabelos cast escuro x loiros	0,0002	14,7512 (1,9888 - 109,4109)
	C/T	cabelos pretos x loiros	0,0002	16,5 (2,1484 - 126,7248)
	T	cabelos pretos x loiros	2,21 x 10⁻⁰⁵	18,8611 (2,5241 - 140,9358)
	C/T	cabelos cast claro x loiros	0,0014	14,2979 (1,8074 - 113,1087)
	T	cabelos cast claro x loiros	0,0021	12,5741 (1,6233 - 97,4009)
	C/C	cabelos loiros x cast escuro	0,0002	16,5161 (2,1896 - 124,5792)
	C	cabelos loiros x cast escuro	0,0002	14,7512 (1,9888 - 109,4109)
	C/C	cabelos loiros x pretos	0,0001	19,6721 (2,5727 - 150,4214)
	C	cabelos loiros x pretos	2,21 x 10⁻⁰⁵	18,8611 (2,5241 - 140,9358)
	C/C	cabelos loiros x cast claro	0,0014	14,2979 (1,8074 - 113,1087)
	C	cabelos loiros x cast claro	0,0021	12,5741 (1,6233 - 97,4009)
	C/C	cabelos loiros x todos	0,0001	16,3687 (2,2202 - 120,6801)
	C	cabelos loiros x todos	0,0001	15,0119 (2,0635 - 109,2093)
rs376784068	T/A	pele escura x todos	0,0270	25,9174 (1,2272 - 547,3528)
	A	pele escura x todos	0,0272	25,5430 (1,2179 - 535,6942)
rs183671	G/G	pele clara x pele escura	3,6 x 10⁻¹⁹	33,6909 (12,4236 - 91,3646)
	G	pele clara x pele escura	6,35 x 10⁻²⁵	12,8778 (7,6662 - 21,6323)
	T/T	pele escura x pele clara	1,41 x 10⁻¹⁰	15,5530 (6,1700 - 39,2054)
	T	pele escura x pele clara	6,35 x 10⁻²⁵	12,8778 (7,6662 - 21,6323)
	T/T	olhos cast escuro x azuis	0,0004	14,1981 (1,8851 - 106,9344)
	T/G	cabelos cast escuro x loiros	3,05 x 10⁻⁰⁶	10,3378 (3,0462 - 35,0835)
	T/T	cabelos pretos x loiros	6,66 x 10⁻⁰⁷	29,2642 (3,8518 - 222,3360)
	T/G	cabelos pretos x loiros	2,96 x 10⁻⁰⁷	13,6667 (3,9433 - 47,3660)
	T	cabelos pretos x loiros	2,39 x 10⁻²²	29,9600 (11,5685 - 77,5904)
	T	cabelos pretos x ruivos	1,55 x 10⁻⁰⁵	10,4256 (2,9690 - 36,6100)
	G/G	cabelos cast claro x pretos	6,52 x 10⁻¹⁰	10,6515 (4,763 - 23,8199)
	G/G	cabelos loiros x cast escuro	1,29 x 10⁻⁰⁸	12,7069 (4,3056 - 37,5016)
	G/G	cabelos loiros x pretos	1,22 x 10⁻¹⁹	67,8333 (20,6051 - 223,3113)
	G	cabelos loiros x pretos	2,39 x 10⁻²²	29,9600 (11,5685 - 77,5904)
	G/G	cabelos loiros x todos	2,69 x 10⁻¹¹	15,3636 (5,3772 - 43,8966)
	G	cabelos loiros x todos	9,58 x 10⁻¹³	11,5352 (4,6171 - 28,8193)
	G/G	cabelos ruivos x pretos	0,0001	16,4444 (3,8176 - 70,8358)
G	cabelos ruivos x pretos	1,55 x 10⁻⁰⁵	10,4256 (2,9690 - 36,6100)	
rs6451048	G/G	pele clara x pele escura	0,0059	24,4286 (1,2930 - 461,5453)
	G	pele clara x pele escura	0,0062	23,5991 (1,2600 - 441,9988)
	A/G	pele escura x pele clara	0,0059	24,4286 (1,2930 - 461,5453)
	A	pele escura x pele clara	0,0062	23,5991 (1,2600 - 441,9988)
	A/G	pele escura x pele média	0,0062	24,0857 (1,2747 - 455,0972)
	A	pele escura x pele média	0,0064	23,2673 (1,2422 - 435,7987)
	A/G	pele escura x todos	0,0007	48,4286 (2,5691 - 912,9152)

	A	pele escura x todos	0,0007	46,8249 (2,5029 - 876,0044)
	G/G	pele média x pele escura	0,0062	24,0857 (1,2747 - 455,0972)
	G	pele média x pele escura	0,0064	23,2673 (1,2422 - 435,7987)
	A/G	cabelos pretos x todos	0,0040	27,5455 (1,4675 - 517,0462)
	A	cabelos pretos x todos	0,0041	26,9466 (1,4433 - 503,1047)
rs6451049	C/C	pele clara x pele escura	1,50 x 10⁻⁰⁷	11,5541 (4,3050 - 31,0099)
	T/C	pele escura x pele clara	1,50 x 10⁻⁰⁷	11,5541 (4,3050 - 31,0099)
rs58784954	G/A	pele escura x todos	0,0151	16,0189 (1,6350 - 156,9412)
	A	pele escura x todos	0,0154	15,6055 (1,6083 - 151,4203)
rs250413	G/A	cabelos pretos x loiros	0,0041	17,8163 (1,0348 - 306,7334)
	G/G	cabelos loiros x pretos	0,0041	17,8163 (1,0348 - 306,7334)
rs551953158	C/T	pele escura x todos	0,0267	26,1009 (1,2359 - 551,2192)
	T	pele escura x todos	0,0269	25,7240 (1,2266 - 539,4858)
rs73077153	C/T	pele escura x todos	0,0267	26,1009 (1,2359 - 551,2192)
	T	pele escura x todos	0,0269	25,7240 (1,2266 - 539,4858)
rs73077154	G/G	pele clara x pele escura	1,39 x 10⁻⁰⁶	21,4651 (4,6606 - 98,8614)
	G	pele clara x pele escura	2,34 x 10⁻⁰⁶	18,7778 (4,1642 - 84,6745)
	G/C	pele escura x pele clara	1,39 x 10⁻⁰⁶	21,4651 (4,6606 - 98,8614)
	C	pele escura x pele clara	2,34 x 10⁻⁰⁶	18,7778 (4,1642 - 84,6745)
	G/C	cabelos pretos x loiros	0,0042	18,1837 (1,0565 - 312,9698)
	G/G	cabelos loiros x pretos	0,0042	18,1837 (1,0565 - 312,9698)
rs143386685	G/A	cabelos pretos x todos	0,0158	21,3353 (1,0908 - 417,3158)
	A	cabelos pretos x todos	0,0160	21,0000 (1,0792 - 408,6275)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Considerando que cada um dos 58 SNPs possui apenas dois alelos, a informação de cada um dos dois torna-se redundante. Portanto, para alelos, foi considerado $\alpha_{\text{corrigido}} = 0,0009$ (isto é, $0,05/58$). Considerando que cada um dos 58 SNPs possui apenas três genótipos, foi considerado $\alpha_{\text{corrigido}} = 0,0003$ (isto é, $0,05/174$).

Tabela 5: Alelos e genótipos de SNPs/Indels do gene *TYR* associados com as características fenotípicas avaliadas.

SNPs/Indels	Alelo/ Genótipo	Característica	Valor de <i>p</i>	Odds Ratio (IC 95%)
rs75741664	C/C	pele clara x pele escura	0,0224	11,0000 (1,2017 - 100,6922)
	C	pele clara x pele escura	0,0231	10,6296 (1,1749 - 96,1677)
	C/T	pele escura x pele clara	0,0224	11,0000 (1,2017 - 100,6922)
	T	pele escura x pele clara	0,0231	10,6296 (1,1749 - 96,1677)
88907815	A/G	cabelos loiros x todos	0,0204	30,6842 (1,4505 - 649,0823)
	G	cabelos loiros x todos	0,0206	30,1813 (1,4380 - 633,4739)
rs142881018	CCTGA/CCTGA	pele clara x pele escura	0,0001	43,7879 (2,4559 - 780,7133)
	CCTGA	pele clara x pele escura	0,0001	41,0190 (2,3225 - 724,4532)
	CCTGA/CCTGA	pele clara x todos	0,0031	17,9164 (1,0470 - 306,5808)
	CCTGA	pele clara x todos	0,0034	17,3932 (1,0207 - 296,3758)
	CCTGA/C	pele escura x pele clara	0,0001	43,7879 (2,4559 - 780,7133)
	C	pele escura x pele clara	0,0001	41,019 (2,3225 - 724,4532)
rs61759512	CCTGA/C	pele escura x todos	0,0005	10,0000 (2,8212 - 35,4454)
	G/G	pele clara x pele escura	0,0232	10,8462 (1,1848 - 99,2924)
	G	pele clara x pele escura	0,0240	10,4815 (1,1585 - 94,8314)
	G/A	pele escura x pele clara	0,0232	10,8462 (1,1848 - 99,2924)
rs11824466	A	pele escura x pele clara	0,0240	10,4815 (1,1585 - 94,8314)
	C/C	pele clara x pele escura	0,0002	23,8333 (2,9057 - 195,4869)
	C	pele clara x pele escura	0,0001	25,0777 (3,1386 - 200,375)
	C	pele clara x todos	0,0033	10,6296 (1,3897 - 81,3041)
	C/T	pele escura x pele clara	0,0006	20,4286 (2,4514 - 170,2374)
	T	pele escura x pele clara	0,0001	25,0777 (3,1386 - 200,375)
	C/T	pele escura x todos	0,0005	10,0000 (2,8212 - 35,4454)

rs55870815	T/C	cabelos pretos x todos	0,0159	21,2515 (1,0865 - 415,6802)
	C	cabelos pretos x todos	0,0161	20,9174 (1,0750 - 407,0224)
rs1042602	A/A	pele clara x pele escura	0,0033	11,0000 (1,4509 - 83,3938)
rs11018530	G/G	pele escura x pele clara	2,69 x 10⁻⁰⁵	50,299 (2,8499 - 887,7312)
	G/G	pele escura x todos	3,89 x 10⁻⁰⁵	15,5556 (3,9854 - 60,7148)
	C/G	olhos cast escuro x azuis	0,0001	10,5326 (2,4366 - 45,5280)
	G	olhos cast escuro x azuis	5,48 x 10⁻⁰⁶	11,3556 (2,7171 - 47,4581)
	C/C	olhos azuis x cast escuro	4,96 x 10⁻⁰⁶	13,1548 (3,0478 - 56,7786)
rs67512950	C	olhos azuis x cast escuro	5,48 x 10⁻⁰⁶	11,3556 (2,7171 - 47,4581)
	A/A	pele escura x pele clara	0,0054	25,2524 (1,3363 - 477,1960)
rs12804012	ACTT/ACTT	olhos azuis x cast escuro	2,58 x 10⁻⁰⁵	11,2778 (2,6121 - 48,6914)
	A/A	pele escura x pele clara	1,67 x 10⁻⁰⁶	64,8065 (3,7250 - 1127,3817)
	A/A	pele escura x todos	1,40 x 10⁻⁰⁶	20,2899 (5,3804 - 76,5139)
	G/G	olhos azuis x cast escuro	3,72 x 10⁻⁰⁶	10,2857 (3,0221 - 35,0073)
rs75375648	G/G	cabelos ruivos x pretos	0,0108	10,0000 (1,2262 - 81,5505)
	T/T	pele clara x pele escura	0,0015	30,6505 (1,6655 - 564,0599)
	T	pele clara x pele escura	0,0017	29,3163 (1,6074 - 534,6844)
	T/C	pele escura x pele clara	0,0015	30,6505 (1,6655 - 564,0599)
rs113438665	C	pele escura x pele clara	0,0017	29,3163 (1,6074 - 534,6844)
	T/T	pele clara x pele escura	0,0015	30,6505 (1,6655 - 564,0599)
	T	pele clara x pele escura	0,0017	29,3163 (1,6074 - 534,6844)
	T/C	pele escura x pele clara	0,0015	30,6505 (1,6655 - 564,0599)
rs1393350	C	pele escura x pele clara	0,0017	29,3163 (1,6074 - 534,6844)
	A/A	olhos cast claro x cast escuro	0,0155	10,8824 (1,2473 - 94,9443)
	A/A	olhos azuis x cast escuro	0,0253	13,0588 (1,3176 - 129,4282)
	A/A	cabelos cast claro x cast escuro	0,0019	16,0741 (1,9303 - 133,8508)
rs1126809	A/A	cabelos cast claro x pretos	0,0091	11,0185 (1,3188 - 92,0605)
	A/A	olhos cast claro x cast escuro	0,0158	10,8209 (1,2399 - 94,4382)
	A/A	olhos azuis x cast escuro	0,0249	13,1818 (1,3288 - 130,7600)
rs28521275	A/A	cabelos cast claro x cast escuro	0,0056	13,4444 (1,5800 - 114,4034)
	C/C	cabelos ruivos x pretos	0,0014	10,8000 (2,6804 - 43,516)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Considerando que cada um dos 42 SNPs possui apenas dois alelos, a informação de cada um dos dois torna-se redundante. Portanto, para alelos, foi considerado $\alpha_{\text{corrigido}} = 0,0012$ (isto é, 0,05/42). Considerando que cada um dos 42 SNPs possui apenas três genótipos, foi considerado $\alpha_{\text{corrigido}} = 0,0004$ (isto é, 0,05/126).

Tabela 6: Haplótipos formados por SNPs/indels das regiões avaliadas para o gene *SLC45A2* associados com as características fenotípicas avaliadas.

Haplótipos	Característica	Valor de p	Odds Ratio (IC 95%)
Promotora			
promo3	cabelos castanhos x loiros	0,0003	12,8864 (1,7513 - 94,8226)
	cabelos pretos x loiros	0,0001	15,8333 (2,1065 - 119,0094)
promo15	cabelos pretos x outros	0,0174	20,2387 (1,0399 - 393,8810)
promo17	pele escura x outras	0,0245	27,2439 (1,2984 - 571,6279)
Codificante isoforma 3			
Iso3cds1	pele clara x pele escura	2,82 x 10⁻²⁵	17,2252 (9,6147 - 30,8598)
	cabelos loiros x castanhos	3,07 x 10⁻⁰⁹	12,6803 (3,9235 - 40,9818)
	cabelos loiros x outros	5,98 x 10⁻¹⁴	18,6224 (5,8151 - 59,6361)
	cabelos loiros x pretos	3,36 x 10⁻²³	50,9371 (15,3526 - 169,0004)
	cabelos ruivos x pretos	1,57 x 10⁻⁰⁵	10,6352 (3,0039 - 37,6531)
Iso3cds2	pele escura x pele clara	4,35 x 10⁻¹⁶	33,4237 (11,3209 - 98,6802)
	pele media x pele clara	7,73 x 10⁻⁰⁸	10,0628 (3,5119 - 28,8329)
	cabelos pretos x loiros	1,09 x 10⁻⁰⁷	29,2778 (3,9315 - 218,0323)
	olhos cast escuro x azuis	0,0001	15,7081 (2,1282 - 115,9398)

	olhos cast escuro x verdes	8,96 x 10⁻⁰⁷	13,3636 (3,1910 - 55,9657)
Iso3cds3	cabelos castanhos x loiros	0,0005	19,5310 (1,1847 - 321,9829)
Iso3cds4	pele escura x pele clara	3,89 x 10⁻⁰⁶	10,2541 (3,5776 - 29,3900)
	cabelos pretos x loiros	0,0002	16,1405 (2,1321 - 122,186)
Iso3cds5	cabelos pretos x outros	0,0128	23,2079 (1,1916 - 452,0176)
Iso3cds7	pele escura x pele clara	0,0138	22,6374 (1,1578 - 442,6151)
Iso3cds11	cabelos pretos x castanhos	0,0268	10,0000 (1,1078 - 90,2664)
	cabelos pretos x outros	0,0121	13,3623 (1,4813 - 120,5401)
Iso3cds13	pele escura x pele clara	0,0032	29,4497 (1,5696 - 552,5619)
	pele escura x pele media	0,0041	27,3195 (1,4558 - 512,6883)
	pele escura x outras	0,0004	56,7160 (3,0262 - 1062,9445)
	cabelos pretos x castanhos	0,0070	22,5162 (1,2043 - 420,9925)
	cabelos pretos x outros	0,0030	30,0542 (1,6082 - 561,6641)
Codificante isoforma 2			
Iso2cds1	cabelos loiros x pretos	6,34 x 10⁻²³	26,1905 (11,6682 - 58,7870)
	cabelos ruivos x pretos	4,28 x 10⁻⁰⁶	10,7143 (3,4205 - 33,5607)
	pele clara x pele escura	3,65 x 10⁻²⁹	22,6306 (12,1594 - 42,1195)
Iso2cds2	cabelos pretos x loiros	9,76 x 10⁻⁰⁸	29,9074 (4,0169 - 222,6729)
	olhos cast escuro x azuis	0,0001	15,3865 (2,0836 - 113,6208)
	olhos cast escuro x verdes	4,52 x 10⁻⁰⁷	13,7295 (3,2788 - 57,4897)
	pele escura x pele clara	4,35 x 10⁻¹⁶	33,4237 (11,3209 - 98,6802)
Iso2cds4	cabelos pretos x loiros	0,0001	16,4876 (2,1784 - 124,7869)
	pele escura x pele clara	3,89 x 10⁻⁰⁶	10,2541 (3,5776 - 29,3900)
Iso2cds6	cabelos pretos x outros	0,0132	22,9068 (1,1761 - 446,1618)
Iso2cds7	pele escura x pele clara	0,0138	22,6374 (1,1578 - 442,6151)
Iso2cds8	olhos cast claro x outros	0,0373	10,5556 (1,0892 - 102,2972)
Iso2cds11	cabelos pretos x outros	0,0125	13,1884 (1,4619 - 118,9742)
Iso2cds12	cabelos pretos x castanhos	0,0074	21,9964 (1,1764 - 411,2910)
	cabelos pretos x outros	0,0031	29,6643 (1,5873 - 554,3880)
	pele escura x pele clara	0,0032	29,4497 (1,5696 - 552,5619)
	pele escura x pele media	0,0044	26,6805 (1,4216 - 500,7262)
	pele escura x outras	0,0004	56,0769 (2,9921 - 1050,9824)
Codificante isoforma 1			
Iso1cds1	cabelos loiros x pretos	1,94 x 10⁻²²	22,9297 (10,765 - 48,8408)
	cabelos ruivos x preto	1,23 x 10⁻⁰⁶	12,2838 (3,9011 - 38,6788)
	pele clara x pele escura	1,08 x 10⁻²⁴	17,3875 (9,4198 - 32,0949)
Iso1cds3	cabelos pretos x loiros	1,93 x 10⁻⁰⁵	32,6962 (1,952 - 547,6619)
	olhos cast escuro x verdes	0,0004	13,6903 (1,8351 - 102,1338)
	pele escura x pele clara	7,08 x 10⁻⁰⁹	55,0000 (7,1113 - 425,3813)
	pele media x pele clara	3,99 x 10⁻⁰⁶	23,5043 (3,1307 - 176,4607)
Iso1cds5	cabelos pretos x loiros	0,0005	14,8487 (1,9519 - 112,9574)
	pele escura x pele clara	3,31 x 10⁻⁰⁶	15,1667 (4,1686 - 55,1813)
Iso1cds6	pele escura x pele clara	4,90 x 10⁻⁰⁵	28,9474 (3,565 - 235,0491)
	pele media x pele clara	0,0044	11,2705 (1,4324 - 88,6796)
Iso1cds7	cabelos pretos x castanhos	0,0005	33,8000 (1,8909 - 604,1793)
	cabelos pretos x outros	0,0001	45,1811 (2,5288 - 807,2252)
Iso1cds9	olhos cast claro x outros	0,0388	10,3577 (1,0688 - 100,3710)
Iso1cds10	pele escura x pele clara	0,0124	23,7485 (1,2142 - 464,5108)
Iso1cds13	cabelos pretos x castanhos	0,0065	23,0520 (1,2327 - 431,0760)
	cabelos pretos x outros	0,0027	30,8141 (1,6486 - 575,9515)

	pele escura x pele clara	0,0028	30,9130 (1,6469 - 580,2418)
	pele escura x pele media	0,0036	28,4534 (1,5156 - 534,1837)
	pele escura x outras	0,0003	59,3106 (3,1634 - 1112,0046)
3'UTR isoforma 2			
Iso2utr1	cabelos loiros x pretos	4,22 x 10⁻¹⁹	17,7273 (8,0766 - 38,9095)
	cabelos ruivos x pretos	3,19 x 10⁻⁰⁶	15,7576 (3,5659 - 69,6328)
	pele clara x pele escura	1,10 x 10⁻²²	11,2414 (6,7414 - 18,7451)
Iso2utr2	cabelos castanhos x loiros	0,0001	14,5500 (1,9824 - 106,7889)
	cabelos pretos x loiros	0,0001	15,9452 (2,122 - 119,8147)
Iso2utr3	olhos cast escuro x verdes	7,20 x 10⁻⁰⁶	19,2188 (2,6116 - 141,4315)
	cabelos pretos x loiros	7,98 x 10⁻⁰⁷	41,0707 (2,4801 - 680,1413)
	olhos cast escuro x azuis	0,0002	22,8947 (1,3908 - 376,8813)
	olhos mel x verdes	0,0405	10,5429 (1,0633 - 104,5386)
	pele escura x pele clara	1,89 x 10⁻⁰⁹	16,7528 (5,6020 - 50,0990)
Iso2utr4	cabelos pretos x loiros	2,09 x 10⁻⁰⁵	19,1268 (2,5595 - 142,9341)
	pele escura x pele clara	8,01 x 10⁻⁰⁹	11,7500 (4,6181 - 29,8962)
Iso2utr5	pele escura x pele clara	1,25 x 10⁻⁰⁶	35,1429 (4,5113 - 273,7634)
	pele media x pele clara	0,0013	13,0455 (1,6847 - 101,0152)
Iso2utr7	pele escura x pele clara	0,0056	24,3803 (1,3016 - 456,6675)
Iso2utr8	cabelos pretos x outros	0,0173	11,7349 (1,3024 - 105,736)
Iso2utr9	pele escura x outras	0,0264	26,0138 (1,2403 - 545,6262)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Para a região promotora foi considerado $\alpha_{\text{corrigido}} = 0,0024$ (isto é, 0,05/21 haplótipos). Para a região codificante da isoforma 3, foi considerado $\alpha_{\text{corrigido}} = 0,0021$ (isto é, 0,05/24 haplótipos). Para a região codificante da isoforma 2, foi considerado $\alpha_{\text{corrigido}} = 0,0018$ (isto é, 0,05/28 haplótipos). Para a região codificante da isoforma 1, foi considerado $\alpha_{\text{corrigido}} = 0,0014$ (isto é, 0,05/36 haplótipos). Para a região 3'UTR da isoforma 2, foi considerado $\alpha_{\text{corrigido}} = 0,005$ (isto é, 0,05/10 haplótipos).

Tabela 7: Haplótipos formados por SNPs/indels das regiões avaliadas para o gene *TYR* associados com as características fenotípicas avaliadas.

Haplótipos	Característica	Valor de p	Odds Ratio (IC 95%)
Promotora			
promo7	pele escura x pele clara	0,0011	33,0000 (1,8072 - 602,5785)
	pele escura x outras	0,0014	14,0659 (2,6881 - 73,6018)
promo8	pele escura x pele clara	0,0001	24,8182 (3,0614 - 201,1940)
	pele escura x pele media	0,0002	21,7273 (2,6789 - 176,2212)
	pele escura x outras	9,95 x 10⁻⁰⁶	23,2727 (4,8616 - 111,407)
promo9	pele escura x pele clara	0,0173	11,8696 (1,3099 - 107,5532)
promo10	pele escura x pele clara	0,0011	33,0000 (1,8072 - 602,5785)
	pele escura x outras	0,0014	14,0659 (2,6881 - 73,6018)
promo15	cabelos pretos x outros	0,0172	20,3531 (1,0453 - 396,2980)
Codificante			
cds5	pele escura x pele clara	0,0002	21,8824 (2,7033 - 177,1274)
	pele escura x pele media	0,0003	20,6275 (2,5478 - 167,0030)
	pele escura x outras	1,71 x 10⁻⁰⁵	21,2549 (4,4492 - 101,5395)
cds7	pele escura x pele clara	0,0061	23,7042 (1,2654 - 444,0331)
cds9	pele escura x pele clara	0,0017	29,2464 (1,6033 - 533,4993)
cds12	ausência x presença sardas	0,0150	21,6144 (1,1105 - 420,7012)

Observação: Os valores em negrito permanecem significativos ao nível de 5% mesmo se a correção de Bonferroni for levada em consideração. Tanto para a região promotora quanto para a codificante foi considerado $\alpha_{\text{corrigido}} = 0,0026$ (isto é, 0,05/19 haplótipos).