Universidade de São Paulo
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Departamento de Química
Programa de Pós-Graduação em Química

**Avaliação da variabilidade genética do gene *MITF* e suas associações com
fenótipos de pigmentação em amostra da população brasileira**

Letícia Marcorin

Dissertação apresentada à Faculdade de
Filosofia, Ciências e Letras de Ribeirão Preto da
Universidade de São Paulo, como parte das exigências para
a obtenção do título de Mestre em Ciências, Área: **Química**

RIBEIRÃO PRETO -SP

2017

LETÍCIA MARCORIN

Avaliação da variabilidade genética do gene *MITF* e suas associações com fenótipos de pigmentação em amostra da população brasileira

Versão corrigida

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Mestre em Ciências.

Área de concentração: Química

Orientador: Prof. Dr. Celso Teixeira Mendes Junior

RIBEIRÃO PRETO -SP

2017

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.

Marcorin, Letícia

Avaliação da variabilidade genética do gene *MITF* e suas associações com fenótipos de pigmentação em amostra da população brasileira. Ribeirão Preto, 2017.

183 p. : il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto/USP – Área de concentração: Química.

Orientador: Mendes-Junior, Celso Teixeira.

1. *MITF*. 2. Pigmentação Humana. 3. Sequenciamento de Nova Geração. 4. Brasil.

Nome: Marcorin, Letícia

Título: Avaliação da variabilidade genética do gene *MITF* e suas associações com fenótipos de pigmentação em amostra da população brasileira

Dissertação apresentada à Faculdade de
Filosofia, Ciências e Letras de Ribeirão Preto da
Universidade de São Paulo, como parte das exigências para
a obtenção do título de Mestre em Ciências, Área: **Química**

Aprovado em:

Banca Examinadora

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Aos meus pais, por todo o apoio e torcida.

Agradecimentos

Primeiramente, agradeço à minha família que sempre apoiou meus estudos, apesar da distância. Mas, em especial aos meus pais, que sempre me incentivaram a estudar e ir atrás das minhas metas, me dando suporte e torcendo pelas minhas conquistas, não só durante meus anos no mestrado, mas toda minha vida. Agradeço imensamente aos dois e a toda família!

Agradeço aos meus amigos, os velhos e os novos. Aos de Santa Bárbara, pela amizade que se manteve mesmo com a distância. Aos de Ribeirão, Anderson, Priscila, Sarah e Lívia, que se tornaram minha família uspiana.

À Itiane, pelas conversas e brindes.

Aos amigos e parceiros de trabalho do Laboratório de Pesquisas Forenses e Genômicas, Alison, Malu, Nádia, os Guilhermes e André pelas discussões que contribuíram para a elaboração desse trabalho e pelas distrações, também tão importantes. Agradeço especialmente ao Alison pela ajuda e parceria na resolução de tantas dúvidas durante o desenvolvimento da nossa metodologia.

Agradeço imensamente ao meu orientador pela oportunidade e pelo aprendizado. Enquanto cursava o bacharelado em química não imaginava que fosse gostar de uma área das ciências biológicas. Contudo, Celso, excelente professor e orientador, me fez mudar de ideia. Muito obrigada!

Agradeço, também, ao Prof. Dr. Erick da Cruz Castelli, pela paciência e disposição para ajudar com as análises de bioinformática sempre que precisei.

Aos professores e à Coordenação do Programa de Pós-Graduação em Química, pelo aprendizado e apoio prestado. Em especial, à Lâmia, pela grande ajuda durante esses anos.

À Universidade de São Paulo, pela infraestrutura.

Ao Prof. Dr. Aguinaldo Luiz Simões, por ceder à utilização dos equipamentos.

À Cláudia, Elisabete, Rubens e Ana Lúcia pelo auxílio prestado nas coletas e na execução de procedimentos importantíssimos para este trabalho.

Ao Laboratório de Imunopatologia Molecular, em especial ao Prof. Dr. Eduardo Antônio Donadi, pela infraestrutura para realização deste trabalho e à Juliana, Sandra e Flávia, pelo auxílio na realização dos ensaios de sequenciamento.

Aos membros da banca examinadora, por aceitarem participar e contribuir com este trabalho.

Finalmente, agradeço a cada voluntário pela disposição e paciência.

Resumo

MARCORIN, L. **Avaliação da variabilidade genética do gene *MITF* e suas associações com fenótipos de pigmentação em amostra da população brasileira.** 2017. 179 páginas. Dissertação de Mestrado - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, 2017.

A cor da pele, olhos e cabelos são alguns dos traços fenotípicos mais aparentes quando nos referimos à identificação de características individuais. Essas características frequentemente são utilizadas na descrição de indivíduos em retratos falados usados em investigações policiais. Porém, em muitos casos as vítimas ou testemunhas não reconhecem o agressor, tornando inviável a produção desses retratos. Contudo, os vestígios biológicos deixados pelo criminoso poderiam ser utilizados na predição de suas características físicas, suprimindo a falta ou complementando o retrato falado. Para que isso seja possível, é preciso conhecer as variáveis responsáveis pela formação desses fenótipos. No caso dos fenótipos de pigmentação há tanto um fator genético, quanto ambiental. Diversos genes participam da formação desses fenótipos, dentre eles está o gene *MITF* (*Melanogenesis-associated transcription factor*), um dos principais regulador da biossíntese de melanina nos melanócitos. Esse gene está fortemente associado às síndromes de Waardenburg e Tietz, as quais causam pigmentação anormal, principalmente na pele, e ao melanoma. No entanto, apesar do claro envolvimento do gene *MITF* na melanogênese, ainda não são conhecidas associações significativas de polimorfismos nesse gene com fenótipos de pigmentação. À vista disso, esse trabalho avaliou a relação da variabilidade do gene *MITF* com os fenótipos de pigmentação encontrados em uma amostra populacional do estado de São Paulo, por meio de sequenciamento de nova geração. Foram identificados 133 pontos de variação em toda a extensão do gene e sua região promotora, dos quais 21 estão associadas a pelo menos um fenótipo de pigmentação de pele, olhos ou cabelo. Adicionalmente foram encontradas associações com ao menos um fenótipo de pigmentação para 3 dos 17 haplótipos da região promotora, 7 dos 50 haplótipos da extensão que engloba a região 5'UTR e codificante, e um dos 18 haplótipos encontrados na região 3'UTR. Considerando os haplótipos encontrados para a extensão total do gene *MITF* e sua promotora, 20 dos 132 haplótipos encontrados estão associados a algum fenótipo de pigmentação. A maior parte das associações encontradas, tanto para alelos e genótipos quanto

para haplótipos, são referentes a fenótipos mais escuros como cabelos castanhos escuros e pretos e pele escura. Associações com fenótipos mais claros, tais como olhos azuis e verdes e cabelos loiros e ruivos, também foram encontradas, porém envolvendo variantes e haplótipos de frequência baixa na população amostrada; tais associações, entretanto, representam achados falsos positivos. Os resultados confirmam a hipótese de que a variabilidade do gene *MITF* pode contribuir para a formação dos fenótipos de pigmentação de pele, olhos e cabelos dos indivíduos da população brasileira.

Palavras-chave: *Microphthalmia associated transcription factor*. Pigmentação humana. Sequenciamento de nova geração.

Abstract

MARCORIN, L. **Evaluation of *MITF* genetic variation and its association with pigmentation phenotypes in a Brazilian population sample.** 2017. 179 pages. Dissertation (Master's Degree) - Faculty of Philosophy, Sciences and Letters at Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo, 2017.

Skin, eye and hair colors are some of the most noticeable phenotypes when referring to the identification of individual characteristics. These characteristics are often used to describe individuals in police sketches used in investigations. However, in many cases the victims or witnesses are unable to recognize the assaulter, making this sketches unfeasible. Nonetheless, biological traces left by the assaulter could be used to predict their physical characteristics, compensating or complementing these sketches. To make this possible, its necessary to know the variables responsible for the development of these traits. Pigmentation phenotype development relies on genetic and environmental aspects. A variety of genes contribute to the development of these phenotypes, among them *MITF* (Melanogenesis-associated transcription factor), one of the main regulators of melanin synthesis in melanocytes. This gene is strongly associated with Waardenburg and Tietz syndrome, which cause abnormal pigmentation, mostly in skin, and melanoma. Although *MITF*'s clear involvement in melanogenesis, significant associations between this gene's polymorphisms and pigmentation phenotypes are still unknown. Thus, this study evaluated the relation between *MITF* genetic variability and pigmentation phenotypes found in a population sample from the state of São Paulo, through next generation sequencing. There were identified 133 variation points through the whole gene and its promoter, from which 21 were associated with at least one skin, eye or hair pigmentation phenotype. Additionally, 3 of the 17 promoter haplotypes, 7 of the 50 haplotypes comprising the 5'UTR and coding regions and one of the 18 3'UTR haplotypes were associated with at least one pigmentation phenotype. Considering the haplotypes found for the whole gene and its promoter, 20 of the 132 haplotypes found were associated with at least one phenotype. The majority of the associations found for alleles, genotypes and haplotypes were related to darker phenotypes, like dark brown and black hair and dark skin. Associations with lighter phenotypes, like blue and green eyes and blonde and red hair, were also found, although involving variants and haplotypes with low frequencies in the studied population; these associations, however,

represent false positives. The results corroborate the hypothesis that the *MITF* variability can contribute to the formation of pigmentation phenotypes in skin, eye and hair in the Brazilian population.

Key words: Microphthalmia associated transcription factor. Human pigmentation. Next generation sequencing.

Lista de Figuras

Figura 1. Papel do gene <i>MITF</i> na regulação da síntese de melanina	20
Figura 2. Dano causado pela radiação UV estimula a expressão do gene <i>MITF</i>	21
Figura 3. Posição dos diferentes inícios de transcrição das isoformas do fator <i>MITF</i>	24
Figura 4. Proporções de indivíduos com diferentes fenótipos de pigmentação de pele, classificação do IBGE, cor de cabelo e cor de olho	43
Figura 5. Distribuição dos 133 pontos de variação em toda a extensão do gene <i>MITF</i>	46
Figura 6. Relação entre os 17 haplótipos encontrados para a região promotora do gene <i>MITF</i>	69
Figura 7. Relação entre os 50 haplótipos encontrados para a região 5'UTR e CD do gene <i>MITF</i>	71
Figura 8. Relação entre os 18 haplótipos encontrados para a região 3'UTR do gene <i>MITF</i> ..	75
Figura 9. Relação entre os 24 haplótipos encontrados para o gene <i>MITF</i> com frequência acima de 1%	78
Figura 10. Representação do desequilíbrio de ligação na região promotora do gene <i>MITF</i> ...	79
Figura 11. Representação do desequilíbrio de ligação na região 5'UTR e CD do gene <i>MITF</i>	80
Figura 12. Representação do desequilíbrio de ligação na região 3'UTR do gene <i>MITF</i>	81
Figura 13. Representação do desequilíbrio de ligação na extensão total do gene <i>MITF</i>	82
Figura 14. Variação do índice de melanina nos indivíduos de pele clara, pele média e pele escura, em áreas expostas e não expostas ao sol.	99
Figura 15. Variação do índice de melanina dentro dos grupos de indivíduos declarados brancos, amarelos, pardos e pretos, em áreas expostas e não expostas ao sol.	99
Figura 16. Distribuição da escala Fitzpatrick dentre os indivíduos auto classificados como brancos, pardos e pretos.....	100
Figura 17. Proporção de indivíduos que possuem <i>singletons</i> dentre o total de indivíduos dentro de cada categoria de cor de cabelo e olhos	121

Lista de Tabelas

Tabela 1. Produção e aproveitamento de clusters e <i>reads</i> em cada sequenciamento	44
Tabela 2. Distribuição de <i>reads</i> de acordo com a qualidade em cada sequenciamento	44
Tabela 3. Qualidade de mapeamento ao genoma de referência.....	45
Tabela 4. Frequências alélicas e genotípicas, heterozigose observada e esperada, e aderência ao Equilíbrio de Hardy-Weinberg dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele	48
Tabela 5. Frequências alélicas e genotípicas, heterozigose observada e esperada, e aderência ao Equilíbrio de Hardy-Weinberg dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos	55
Tabela 6. Frequências alélicas e genotípicas, heterozigose observada e esperada, e aderência ao Equilíbrio de Hardy-Weinberg dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos.....	62
Tabela 7. Haplótipos encontrados para região promotora.....	70
Tabela 8. Haplótipos encontrados para região 5'UTR e codificante	72
Tabela 9. Haplótipos encontrados para região 3'UTR.....	76
Tabela 10. Haplótipos encontrados para a extensão total do gene <i>MITF</i> com frequência mínima de 1%	77
Tabela 11. Associações fenotípicas encontradas para 21 pontos de variação dentre os 133 analisados em toda a extensão do gene <i>MITF</i>	83
Tabela 12. Associações fenotípicas encontradas para três haplótipos dentre os 17 encontrados para a região promotora.....	87
Tabela 13. Associações fenotípicas encontradas para sete haplótipos dentre os 51 encontrados para a região 5'UTR e CD.....	87
Tabela 14. Associações fenotípicas encontradas para um único haplótipo dentre os 18 encontrados para a região 3'UTR.....	88
Tabela 15. Associações fenotípicas encontradas para 20 haplótipos dentre os 132 encontrados para a extensão total do gene <i>MITF</i>	88
Tabela 16. Fatores de Transcrição que podem ter sua eficiência de ligação ao gene <i>MITF</i> alterada pelas variantes associadas à algum fenótipo de pigmentação presentes em regiões regulatórias	91
Tabela 17. Valores de estabilidade para os 20 miRNAs com maior ΔE dentre os 150 mais estáveis	95
Tabela 18. Valores de especificidade para os 20 miRNAs com maior ΔS dentre os 150 mais específicos	96
Tabela 19. Valores de ES para os 20 miRNAs com maior ΔES dentre os 150 de maior especificidade e estabilidade	97

Tabela 20. Distribuição dos 50 miRNAs que apresentaram os maiores valores de ΔE , ΔS e ΔES nas interações com os diferentes haplótipos encontrados para 3'UTR do gene <i>MITF</i>	98
Tabela 21. Comparação do índice de melanina considerando 17 pontos de variação associados a algum fenótipo de pigmentação (<i>singletons</i> não estão inclusos)	102
Tabela 22. Frequências alélicas das variantes da região promotora do gene <i>MITF</i> nas populações mundiais.	113
Tabela 23. Frequências alélicas das variantes da região codificante do gene <i>MITF</i> nas populações mundiais.	117
Tabela 24. Frequências alélicas das variantes da região 3'UTR do gene <i>MITF</i> nas populações mundiais.	119
Tabela 25. Frequências alélicas das variantes com associações fracas a fenótipos de pigmentação no gene <i>MITF</i> nas populações mundiais.....	124

Sumário

1. <u>INTRODUÇÃO</u>	16
1.1. IDENTIFICAÇÃO HUMANA	17
1.2. MELANOGÊNESE	19
1.3. PIGMENTAÇÃO HUMANA	22
1.4. <i>MITF</i>	23
2. <u>JUSTIFICATIVA</u>	27
3. <u>HIPÓTESE</u>	29
4. <u>OBJETIVOS</u>	31
5. <u>MATERIAL E MÉTODOS</u>	33
5.1. COLETA	34
5.2. PROCEDIMENTOS EXPERIMENTAIS	35
5.2.1. EXTRAÇÃO	35
5.2.2. QUANTIFICAÇÃO	36
5.2.3. PREPARO DE BIBLIOTECAS	36
5.2.4. SEQUENCIAMENTO	37
5.3. PROCESSAMENTO DE DADOS DO SEQUENCIAMENTO	37
5.3.1. TRIMAGEM E OBTENÇÃO DOS GENÓTIPOS	37
5.4. OBTENÇÃO DOS HAPLÓTIPOS	38
5.5. ANÁLISES ESTATÍSTICAS	39
6. <u>RESULTADOS</u>	41
6.1. AMOSTRA POPULACIONAL	42
6.2. SEQUENCIAMENTO E TRATAMENTO COMPUTACIONAL	43
6.3. <i>VARIANT CALLING</i>	45
6.4. FREQUÊNCIAS ALÉLICAS, GENOTÍPICAS, HAPLOTÍPICAS E HETEROZIGOSE	46
6.4.1. PROMOTORA	69
6.4.2. 5'UTR E CD	71
6.4.3. 3'UTR	75
6.4.4. <i>MITF</i>	76
6.5. DESEQUILÍBRIO DE LIGAÇÃO	79
6.6. ASSOCIAÇÕES FENOTÍPICAS	83

6.7. INTERAÇÕES TF/TFBS.....	90
6.8. INTERAÇÃO MIRNA/MRNA.....	94
6.9. RELAÇÃO ENTRE ÍNDICE DE MELANINA E OS FENÓTIPOS E GENÓTIPOS.....	99
7. <u>DISCUSSÃO</u>	104
7.1. AMOSTRA POPULACIONAL.....	105
7.2. <i>VARIANT CALLING</i>	106
7.3. DESVIOS DO EQUILÍBRIO DE HARDY-WEINBERG	107
7.4. ASSOCIAÇÕES FENOTÍPICAS	108
7.4.1. PROMOTORA	109
7.4.2. 5'UTR E CD	113
7.4.3. 3'UTR.....	117
7.4.3.1. INTERAÇÕES MIRNA/MRNA	119
7.4.4. <i>SINGLETONS</i>	120
7.4.5. ASSOCIAÇÕES ENCONTRADAS PARA A EXTENSÃO COMPLETA DO GENE <i>MITF</i>	121
8. <u>CONCLUSÕES</u>.....	125
<u>REFERÊNCIAS</u>	127
<u>APÊNDICE A – VARIANTES DO GENE MITF</u>	135
<u>APÊNDICE B - TABELA 10 COMPLETA: HAPLÓTIPOS ENCONTRADOS PARA A EXTENSÃO TOTAL DO GENE MITF</u>.....	139
<u>ANEXO A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO</u>	143
<u>ANEXO B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA ARMAZENAMENTO DE MATERIAL BIOLÓGICO</u>	145
<u>ANEXO C - PROTOCOLO DE ANÁLISE LABORATORIAL (HALOPLEX) PARA SEQUENCIAMENTO DE NOVA GERAÇÃO</u>	147
<u>ANEXO D - PROTOCOLO DE PROCESSAMENTO COMPUTACIONAL DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO</u>.....	155
<u>ANEXO E - PROTOCOLO DE ANÁLISE ESTATÍSTICA DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO</u>	171
<u>ANEXO F - PROTOCOLO PARA IDENTIFICAÇÃO COMPUTACIONAL DE INTERAÇÕES ENTRE MIRNAS E REGIÃO 3'UTR</u>	180

1. Introdução

1. Introdução

1.1. Identificação humana

Metodologias eficientes que permitem identificar ou diferenciar indivíduos são extremamente importantes em situações como desastres em massa e na elucidação de crimes. Nos casos envolvendo uma quantidade grande de vítimas e em que os corpos não podem ser facilmente reconhecidos, geralmente são utilizados registros odontológicos para a comparação da arcada dentária, pois essa metodologia é simples, eficaz e de baixo custo (Velho et al. 2011). Por outro lado, essa técnica é pouco utilizada na identificação de autores de crimes, situação na qual não existe material de partida para comparação. Nesses casos, costuma-se utilizar a papiloscopia (impressões digitais) ou exame de DNA. Essas duas metodologias comparam materiais coletados na cena do crime com o material de algum suspeito e, por isso, são ineficientes nos casos em que a investigação não consegue levantar suspeitos. No Trabalho “Investigação Criminal de Homicídios”, publicado pela Secretaria Nacional de Segurança Pública (SENASP) em 2014, consta que materiais biológicos são utilizados mais no sentido de comprovar suspeitas do que propriamente desvendar a autoria dos homicídios e, por isso, em casos em que a investigação policial não consegue levantar suspeitos, tais materiais não chegam a ser processados (Maria e Luca 2014). Porém, com a implantação da Rede Integrada de Bancos de Perfis Genéticos (RIBPG), formalizada em 2013 pelo decreto nº 7.950, todos os vestígios biológicos encontrados em locais de crime devem ser coletados e os perfis genéticos obtidos devem ser cadastrados na RIBPG, assim como o perfil genético de presos por crimes hediondos.

Contudo, o material biológico encontrado em cenas de crime poderia ser utilizado para identificar características fenotípicas como a cor da pele, olhos e cabelo além de outros traços biogeográficos, auxiliando a investigação policial. Já estão sendo desenvolvidas algumas metodologias para caracterizar traços faciais, ancestralidade e pigmentação (Claes, Liberton, et al. 2014; Claes, Hill, et al. 2014; Walsh et al. 2013; Walsh et al. 2011). Contudo, apesar da cor da pele, olhos e cabelo ser bastante predizível, a diversidade dessas características devido a fatores genéticos, como a ancestralidade, e ambientais, como a exposição a luz solar, torna essa predição um tanto complicada dependendo da população estudada (Balaesque e King 2016). Assim, a população brasileira é especialmente complexa devido a miscigenação (Pena et al.

2011) e alta incidência solar em grande parte do país. Portanto, é importante conhecer a variabilidade genética dos genes envolvidos na produção de melanina, assim como a composição ancestral da população estudada para que seja possível prever características fenotípicas com maior acurácia.

Outro desafio é o desenvolvimento de uma metodologia rápida e barata que possa ser aplicada na identificação dessas características. Atualmente a genotipagem de SNPs (*Single Nucleotide Polymorphism*) é um tanto mais trabalhosa e cara que a de STRs (*Short Tandem Repeats*) utilizada pelos departamentos de polícia científica brasileiros. A genotipagem dos STRs utilizados pode ser feita de maneira simultânea com kits multiplex disponíveis atualmente, como o PowerPlex® 16 System da Promega (Bonaccorso 2005), que permite genotipar 15 STRs e amelogenina (cromossomos sexuais). A análise multiplex dos SNPs pode ser feita por meio de kits como, por exemplo, o *SNaPshot® Multiplex Kit* (Applied Biosystems) que permite a genotipagem de até 10 SNPs por reação. Ao contrário dos STRs que têm um alto poder discriminatório, a genotipagem de SNPs para fins de identificação humana necessita avaliar uma quantidade bem maior de pontos de variação para alcançar o nível de informatividade dos STRs (Musgrave-Brown et al. 2007).

Esta limitação metodológica dificulta o emprego de SNPs na prática forense. No que se refere à predição de características morfológicas, o sistema HIrisPlex, sozinho, une 24 SNPs que predizem a cor dos olhos e cabelos (Walsh et al. 2011). Quanto a predição da cor da pele, Maroñas et al estudaram 59 SNPs dos quais 29 mostraram relação com a pigmentação da pele (Maroñas et al. 2014). A predição de ancestralidade também pode trazer informações relevantes quanto ao fenótipo. Pachou et al propôs 5.000 marcadores altamente informativos para a estimativa de ancestralidade (Paschou et al. 2010). Além disso, um estudo conduzido por Claes et al. encontrou 24 SNPs que trazem informações sobre características faciais (Claes, Hill, et al. 2014). Contudo, mesmo a somatória desses SNPs não é suficiente para inferir com confiança as características fenotípicas, principalmente em relação à cor da pele e características faciais (Kayser 2015). Portanto, o número de SNPs necessários para o *DNA phenotyping* torna a técnica significativamente mais cara e trabalhosa.

Entretanto, os avanços tecnológicos nessa área caminham para a redução do custo por amostra analisada. As técnicas de sequenciamento de nova geração (*Next Generation Sequencing*, NGS) permitem o sequenciamento de um número cada vez maior de regiões do genoma de vários indivíduos de uma só vez, com um menor custo por amostra sequenciada. Desde o sequenciamento do primeiro genoma humano até hoje os custos caíram de bilhões para algumas centenas de dólares. Além da diminuição dos custos, as tecnologias futuras visam

reduzir, também, a complexidade e quantidade de reagentes necessários para o preparo de amostras, tornando o sequenciamento de DNA cada vez mais vantajoso (Levy e Myers 2016; Yang et al. 2014).

1.2. Melanogênese

Durante a embriogênese, os melanoblastos migram da crista neural para a pele, olhos e folículos capilares, se diferenciando em melanócitos. Essa diferenciação é regulada por diversos processos, dentre eles a ativação do gene *MITF* (*Melanogenesis-associated transcription factor* ou *Microphthalmia-associated transcription factor*) a partir da interação do hormônio estimulante do melanócito (*alpha-Melanocyte Stimulating Hormone*, α -MSH) com a proteína transmembrana MC1R (*Melanocortin 1 receptor*). O fator MITF regula a expressão das proteínas TYR (tirosinase), TYRP1 (*Tyrosinase related protein 1*), DCT (*Dopachrome tautomerase*), entre outras, permitindo a diferenciação do melanoblasto em melanócito (Hirobe 2011).

A produção de pigmento ocorre dentro dos melanócitos em vesículas chamadas melanossomos e é intermediada por algumas enzimas que convertem a tirosina em melanina (Sturm e Frudakis 2004). A síntese de melanina pode tomar dois caminhos reacionais que levam a produção de um dos dois pigmentos: feomelanina (pigmento vermelho-amarelo) ou eumelanina (pigmento marrom-preto) (Figura 1). O funcionamento basal do melanossomo resulta na produção de feomelanina e depende da presença de cisteína e da enzima tirosinase, que é altamente expressa nos melanócitos. Nesse momento essas organelas se diferenciam em feomelanossomos, que são menores, e têm uma estrutura mais frágil (Sturm e Frudakis 2004). Já nos eumelanossomos ocorre a produção de eumelanina, que depende da presença de outras enzimas como a TYRP1 e DCT, além da tirosinase (Fang et al. 2002; Cheli et al. 2010).

Figura 1. Papel do gene *MITF* na regulação da síntese de eumelanina e feomelanina nos melanócitos. A proteína codificada pelo gene *MITF* regula a expressão das proteínas transmembrana TYR, TYRP1 e DCT que estão diretamente envolvidas na síntese dos pigmentos, além de regular também a expressão do gene *OCA2* que transcreve a proteína transmembrana P, possivelmente responsável pelo controle do pH no melanossomo (Tully 2007; Sturm 2009; Ancans et al. 2001). A estrutura que representa a eumelanina nessa figura é uma das várias estruturas de monômeros que podem compor esse polímero (Riesz 2007). Figura criada com o auxílio da ferramenta ACD/ChemSketch (ACD/Labs Inc.).

Os melanossomos maduros, contendo pigmento, são transferidos para os queratinócitos na epiderme, onde se acumulam preferencialmente próximos ao núcleo. A presença dos melanossomos auxilia na proteção da célula contra a radiação UV, agindo como um protetor solar natural (Maddodi et al. 2012). Os pigmentos possuem estruturas bastante conjugadas permitindo a absorção de um espectro amplo de radiação, abrangendo do UV (100nm a 400nm, aproximadamente) ao visível (400nm a 780nm, aproximadamente). Geralmente, o aumento no número de conjugações (isto é, o aumento de ‘ligações duplas intercaladas’) desloca a absorção para frequências mais baixas, menos energéticas, de maior comprimento de onda. Isso explica a absorção, principalmente da feomelanina, no espectro de luz visível. Já a absorção da radiação UV, feita principalmente pela eumelanina, é explicada pela variedade de estruturas distintas formadas pela polimerização de diferentes proporções dos monômeros DHI (5,6-dihidroxiindol, de coloração preta) e DHICA (ácido 5,6-dihidroxiindol-2-carboxílico, de coloração marrom), dois precursores da eumelanina (Figura 1). As proporções dos dois monômeros na estrutura da eumelanina depende de alguns fatores, dentre eles o pH do meio reacional e a presença da enzima DCT. A decomposição espontânea da dopacroma (DOPA, Figura 1) produz DHI e CO_2 . Essa reação é menos espontânea dentro do melanossomo, onde o

pH levemente ácido desfavorece a saída do grupo carboxila ($-\text{COOH}$). Por isso, DHI costuma ser encontrado em menor proporção. Já a formação de DHICA depende da ação da enzima DCT, que catalisa a tautomerização da DOPA. A somatória dos comprimentos de onda absorvidos pelas estruturas oligoméricas formadas através da polimerização de diferentes proporções dos monômeros DHI e DHICA resultaria em um espectro de absorção bastante amplo abrangendo do UV ao visível. A energia absorvida por essas moléculas se distribui por sua estrutura e é dissipada sem danificá-la, na forma de calor, por exemplo (Pezzella et al. 1996; Pezzella et al. 2003; Riesz 2007).

Como descrito por D'Orazio et al., a exposição à radiação solar estimula a produção de pigmento como uma resposta ao dano causado à célula: o stress oxidativo causado pela radiação aumenta a produção de espécies reativas de oxigênio que se acumulam no queratinócito e ativam a cadeia de resposta ao dano mediada pela proteína p53, que por sua vez ativa a expressão do gene POMC (*Proopiomelanocortin*), aumentando a síntese de α -MSH que acaba por estimular a expressão do gene *MITF* através da sua interação com a proteína transmembrana MC1R nos melanócitos, intensificando a produção de eumelanina (Figura 2) (D'Orazio et al. 2013).

Figura 2. Dano causado pela radiação UV estimula a expressão do gene *MITF* pela ativação da p53, intensificando a produção de eumelanina nos melanócitos (Adaptada de D'Orazio et al., 2013).

Além da influência da radiação UV, a produção de pigmento depende ainda do pH do meio reacional. Arcans et al. descreveram que a tirosinase tem atividade ótima em pH 6,8 e, portanto, o pH ácido nos melanossomos acarretaria em uma produção mais baixa de pigmento. Nesse mesmo estudo, os pesquisadores propõem que a proteína transmembrana P, expressa pelo gene *OCA2*, também regulado pelo gene *MITF*, funcionaria como um neutralizador de pH, proporcionando um meio mais favorável para a atividade da tirosinase e intensificando a produção de pigmento (Ancans et al. 2001).

1.3. Pigmentação humana

Apesar da radiação solar ser prejudicial para a pele em altas intensidades, ela é essencial para a produção de vitamina D pelo nosso organismo. A vitamina D regula a absorção de cálcio, sendo necessária para o desenvolvimento e manutenção dos ossos (Rossberg et al. 2016). Estudos mostram que indivíduos com pele clara produzem vitamina D com maior facilidade, uma vez que a radiação passa com mais facilidade pela pele (Bonilla et al. 2014). A de-pigmentação da pele dos indivíduos europeus pode ter sido, dentre outros fatores, uma adaptação à irradiação solar menos intensa nas regiões do hemisfério norte, o que dificultaria a produção de vitamina D em indivíduos de pele mais escura (Jablonski e Chaplin 2010).

De fato, povos que vivem em regiões de altas latitudes, como os europeus e asiáticos, geralmente têm pele mais clara, enquanto povos que vivem em baixas latitudes, como grande parte dos africanos, possuem pele mais escura. No entanto, populações mais miscigenadas, como as do continente americano, fogem desse padrão, contendo indivíduos de diferentes tons de pele em regiões de altas e baixas latitudes. Por conta disso, as características de pigmentação podem não representar confiavelmente a ancestralidade do indivíduo (Ruiz-Linares et al. 2014).

A população brasileira é bastante diversificada, devido à grande quantidade de imigrantes, principalmente dos continentes europeu e africano. Como consequência, as contribuições europeia e africana na ancestralidade da população brasileira são bastante acentuadas. Segundo o censo demográfico de 2010 (Instituto Brasileiro de Geografia e Estatística - IBGE), a população brasileira é composta por 47,7% de brancos, 7,6% de indivíduos declarados pretos, 43,1% declarados pardos, 0,4% declarados indígenas e 1,1% amarelos. Estatísticas sobre as características étnico-raciais da população brasileira em 2008

(IBGE) mostraram que mais da metade da população (51,6%) se identifica com pelo menos duas raças ou cor de pele.

A região sudeste do país, mais especificamente a região de Ribeirão Preto, no estado de São Paulo, segundo dados do censo de 2010 (IBGE), é composta por 69,8% indivíduos declarados brancos, 6,4% pretos, 22,8% pardos, 0,9% amarelos e 0,1% indígenas. A diferença entre os dados apresentados para o país como um todo e para a região de Ribeirão Preto, mostra que é necessária a caracterização da população em cada região do país que, por ser bastante extenso, possui populações com características diferentes ao longo de seu território.

1.4. *MITF*

O gene *MITF* (*Melanogenesis associated transcription factor*, antigamente chamado de *Microphthalmia-associated transcription factor*) está localizado na banda 13 do braço curto do cromossomo 3 e transcreve proteínas com estruturas hélice-alça-hélice (*basic helix-loop-helix*) e zíper de leucina. Essas proteínas formam dímeros (com outras proteínas de estrutura semelhante, ou homodímeros que atuam como fatores de transcrição na regulação de diferentes genes envolvidos no desenvolvimento do melanócito e na síntese de pigmento nos melanossomos (Lu et al. 2010). Além disso, mutações no gene *MITF* estão associadas ao desenvolvimento de síndrome de Waardenburg tipo 2 (Takeda et al. 2000; Zhang et al. 2012; Grill et al. 2013) e síndrome de Tietz (Smith et al. 2000; Tachibana 2000; Grill et al. 2013), caracterizadas por pigmentação anormal da pele, olhos e cabelo, surdez, e melanoma (D’Orazio et al. 2013; Hartman e Czyz 2015; Haflidadóttir et al. 2010).

Existem pelo menos 8 isoformas conhecidas do fator MITF (A, B, C, D, E, H, J, M) que possuem diferentes regiões promotoras podendo gerar diferentes mRNAs e/ou transcritos, mas compartilham a porção contendo as estruturas de hélice-alça-hélice e zíper de leucina, essenciais para a interação do fator com as regiões dos genes regulados por ele (Primot et al. 2010) (Figura 3). Essas isoformas costumam ser tecido-específicas: isoformas A, B, D e H são expressas no olho (*retinal pigment epithelium*, RPE) (Udono et al. 2000; Bharti et al. 2008; Maruotti et al. 2012); As isoformas E, M e C foram encontradas em mastócitos (Oboki et al. 2002; Takemoto et al. 2002); Já MITF-J é expresso em vários tecidos (Hershey e Fisher 2005), assim como a isoforma C (Fuse et al. 1999). A isoforma M é expressa quase que exclusivamente em melanócitos (Tachibana 2000), mas também é encontrada na retina (Maruotti et al. 2012).

MITF-M não contém a porção 1B1b do éxon B (Figura 3), que é comum nas outras isoformas, fazendo com que esse fator se mantenha no núcleo, uma vez que a porção 1B1b auxilia no transporte do fator para o citoplasma. Como consequência disso, o fator de transcrição MITF-M tem atividade intensificada nos melanócitos (Lu et al. 2010). Uma nona isoforma, Mitf-Mc, foi identificada em ratos, exclusivamente em mastócitos (Takemoto et al. 2002). Apesar do gene humano possuir o ponto de iniciação homologa para a transcrição da isoforma MITF-Mc, ainda não foram encontrados mRNAs correspondentes à essa isoforma, inclusive em mastócitos (Hershey e Fisher 2005).

Figura 3. Ilustração da posição dos diferentes inícios de transcrição das isoformas do fator MITF (A) e exemplificação das estruturas dos transcritos que codificam essas isoformas (B). Todos os transcritos compartilham os éxons 2 a 9 que determinam as estruturas de hélice-alça-hélice e zíper de leucina necessárias para a atividade do fator. O códon de início de tradução (ATG) em cada transcrito está indicado com um asterisco (Adaptado de Hershey e Fisher 2005).

Durante o desenvolvimento celular, o fator de transcrição MITF estimula a expressão dos genes *TYR*, *TYRP1* e *DCT*, fazendo com que o melanoblasto se diferencie em melanócito e migre para a região basal da epiderme. A partir daí, a regulação da expressão dos genes *TYR*, *TYRP1* e *DCT* pelo gene *MITF* estimulará a produção de pigmento. Essa regulação acontece a partir da interação do fator de transcrição MITF com a região E (*E-box*, '(T)CATGT(G)'), ou uma versão estendida dessa sequência chamada de região M (M-box, AGTCATGTGCT), localizadas na região promotora próxima ao início da transcrição dos genes *TYR*, *TYRP1* e *DCT* (Vachtenheim e Borovanský 2010). O fator MITF ainda regula a transcrição dos genes *OCA2*

e *BCL2* que também possuem uma versão de E-box (Vachtenheim e Borovanský 2010). O gene *OCA2* expressa a proteína transmembrana P, dita responsável por regular o pH no melanossomo para que a atividade da TYR seja ótima (Ancans et al. 2001; Tully 2007; Sturm 2009). Já o gene *BCL2* está envolvido na perda de pigmentação dos cabelos, podendo ser responsável pela perda de melanócitos no bulbo capilar, uma vez que esse gene expressa uma proteína com ação anti-apoptótica (Nishimura 2005; Mak et al. 2006).

A expressão do fator MITF-M é regulada por diferentes vias, sendo a principal a desencadeada pela interação do hormônio estimulante do melanócito (α -MSH) com a proteína transmembrana MC1R nos melanócitos. Essa interação acarreta em um aumento da concentração de AMP cíclico (cAMP) e ativa a proteína quinase A (PKA) que vai, então, ativar CREB (cAMP response element binding protein). CREB se liga a promotora do gene *MITF* (posições -140 e -147) intensificando sua expressão (Wan et al. 2011) (Figura 1). Como descrito anteriormente, essa via geralmente é desencadeada pela ativação do p53 nos queratinócitos em resposta ao dano causado pela radiação solar (Figura 2). Nesse caso há um aumento na expressão do gene *MITF* e, conseqüentemente, dos genes *TYR*, *TYRP1*, *DCT* e *OCA2* o que ocasiona um aumento na produção de pigmento dentro dos melanossomos que serão transportados para os queratinócitos (Vachtenheim e Borovanský 2010). A proteína expressa pelo gene *ASIP* (Agouti signaling protein), assim como α -MSH, pode se ligar a proteína transmembrana MC1R bloqueando a interação dessa proteína com α -MSH, diminuindo a produção de pigmento escuro (eumelanina). A expressão elevada de *ASIP* leva a fenótipos mais claros (Nan et al. 2009).

Além dessa via, a expressão do gene *MITF* é modulada por alguns outros fatores de transcrição: SOX10 (Sry-related HMG box 10), além de ativar a transcrição de MITF-M, ainda trabalha sinergicamente com o fator para ativar a expressão de *DCT*; PAX3 (Paired box 3) pode agir sozinho ou junto com SOX10 na ativação da expressão de MITF-M, além de competir com esse fator pelos sítios de ativação de *TYR*, *TYRP1* e *DCT*, impedindo as expressões desses genes; estudos mostram que STAT3 (*Signal transducer and activator of transcription 3*) pode intensificar a expressão do gene *MITF*, mas essa relação ainda não é bem compreendida (Wan et al. 2011).

Existem vários estudos relacionando o envolvimento do gene *MITF* no desenvolvimento de doenças como síndromes que causam a pigmentação anormal, principalmente da pele, e melanoma. O envolvimento do gene nas síndromes de Waardenburg e Tietz foi caracterizado por 21 mutações presentes nos éxons que codificam a porção hélice-alça-hélice e zíper de leucina do fator MITF, além de regiões regulatórias de *splicing* (Pingault et al. 2010).

Recentemente, um estudo descreveu duas mutações não sinônimas no gene (uma delas já descrita como causadora de síndrome de Waardenburg) que, juntas, causam uma síndrome chamada pelo grupo de COMMAD (*Coloboma, Osteopetrosis, Microphthalmia, Macrocephaly, Albinism, and Deafness*) que tem como sintomas a coloboma e microftalmia, que causam má formação ocular, osteopetrose, caracterizada pelo desenvolvimento de ossos densos e quebradiços, microcefalia (má formação craniana e cerebral), além de albinismo e surdez, sintomas causados pela síndrome de Waardenburg (George et al. 2016). Além dessas doenças, diferentes níveis de expressão do gene levam ao desenvolvimento e proliferação de melanoma, porém mutações que afetam a expressão do gene *MITF* ocorrem em apenas aproximadamente 10% dos casos da doença (Koludrovic e Davidson 2013; Cronin et al. 2009). Uma das mutações mais frequentemente associadas ao melanoma, rs149617956 (c.G1075A), foi também, associada a olhos não azuis em estudo que avaliou a relação da presença dessa variante com a diferença na suscetibilidade de desenvolvimento de melanoma entre indivíduos com fenótipos claros e escuros em amostras populacionais da Austrália e Inglaterra (Yokoyama et al. 2011). Entretanto, a variabilidade do gene *MITF* com relação a fenótipos de pigmentação ainda não foi estudada de maneira compreensiva e atualmente se conhece muito pouco da relação desse gene com a variação normal da pigmentação humana (Kayser 2015).

2. Justificativa

2. Justificativa

Compreender a relação entre os processos moleculares que determinam características fenotípicas e os polimorfismos genéticos associados à pigmentação humana beneficiaria, principalmente, duas áreas da ciência. A medicina personalizada poderia utilizar essas informações para entender melhor os processos que levam ao desenvolvimento de cânceres e outras doenças caracterizadas pela pigmentação anormal, principalmente da pele. Mutações no gene *MITF*, por exemplo, podem levar ao desenvolvimento de melanoma (Hartman e Czyz 2015; Lauss et al. 2015), um câncer muito agressivo, síndrome de Waardenburg e de Tietz que causam pigmentação anormal da pele e surdez (Grill et al. 2013; Zhang et al. 2012; Smith et al. 2000). Portanto o estudo do funcionamento e regulação desse gene pode contribuir para o desenvolvimento de tratamentos mais eficazes para esses distúrbios, uma vez que os tratamentos utilizados atualmente são pouco eficientes e ou invasivos. A utilização de miRNAs parece ser uma boa alternativa para tratamentos dermatológicos, por exemplo (Konda et al. 2012; Osborne et al. 2012; Ning e Andl 2013).

Já a área forense poderia se beneficiar utilizando marcadores que estejam relacionados às características fenotípicas, que poderiam ser genotipados em amostras biológicas encontradas em locais de crime, para obter informações sobre características físicas de indivíduos envolvidos no tal crime. Essa metodologia, chamada Forensic DNA phenotyping, já é estudada por alguns grupos de pesquisa (Claes, Hill et al. 2014; Claes, Liberton et al. 2014; Walsh et al. 2013; Kayser 2015). A obtenção de informações sobre as características físicas de um indivíduo, como a cor da pele, olhos e cabelos, pode ser muito útil em casos em que não existem suspeitos, direcionando as investigações. Como o gene *MITF*, foco deste estudo, é um dos principais reguladores do processo de melanogênese (Liu e Fisher 2010), é indispensável que se conheça sua diversidade e a influência desta em seu funcionamento e regulação, para que seja possível relacionar um determinado genótipo ou haplótipo a um fenótipo específico. Além disso, o fato de a relação genótipo-fenótipo ser dependente da população, de sua composição ancestral e localização geográfica (Balaesque e King 2016), torna importante que esse estudo seja feito em grupos populacionais brasileiros.

3. Hipótese

3. **Hipótese**

Variantes presentes na extensão do gene *MITF* e os haplótipos que compreendem essas variantes contribuem para a formação dos diferentes fenótipos de pigmentação humana, uma vez que o produto desse gene atua na regulação da expressão de diversos outros genes (*TYR*, *TYRP1*, *DCT*, *OCA2*, *BCL2*, por exemplo) imprescindíveis para a biossíntese de melanina.

4. Objetivos

4. Objetivos

Analisar, por meio de sequenciamento de nova geração, a diversidade genética das regiões promotora, exônicas e 3'UTR do gene *MITF*, que está envolvido no processo de biossíntese de melanina em amostras de 340 indivíduos da população do estado de São Paulo. Isso permitirá:

- Determinar as frequências alélicas e haplotípicas dessas regiões.
- Caracterizar as regiões regulatórias, promotora e 3'UTR, do gene *MITF*.
- Caracterizar a região codificante do gene *MITF*, buscando avaliar a diversidade de variantes proteicas por ele codificadas.
- Avaliar a intensidade do desequilíbrio de ligação (LD) dentro das regiões regulatórias e exônicas analisadas no gene *MITF*.
- Identificar fatores de transcrição que atuam como reguladores da expressão do gene *MITF* e avaliar o impacto de variantes presentes nas regiões regulatórias nesta interação.
- Identificar miRNAs que atuam como reguladores pós-transcricionais do gene *MITF* e avaliar o impacto de variantes presentes na região 3'UTR nesta interação.
- Identificar polimorfismos presentes nas regiões exônicas do gene *MITF* que possam alterar a proteína MITF produzida.
- Identificar polimorfismos distribuídos pelas regiões promotora, codificante e 3'UTR que possam contribuir para a determinação da pigmentação da pele, cabelos e olhos, e presença de sardas, em indivíduos da população brasileira.

5. Material e Métodos

5. Material e Métodos

5.1. Coleta

Foram amostrados 346 indivíduos de sexo feminino (168) e masculino (178) com idades entre 18 e 70, abrangendo as diferentes características de pigmentação estudadas nesse trabalho. Do total, 206 indivíduos foram amostrados aleatoriamente enquanto que 140 foram priorizados visando enriquecer a amostra com fenótipos mais raros.

Conforme proposto no projeto aprovado pelo Comitê de Ética e Pesquisa da FFCLRP/USP (Processo CEP – FFCLRP – CAAE n.25696413.7.0000.5407), todos os doadores assinaram os devidos Termos de Consentimento Livre e Esclarecido (Anexos A e B) e receberam uma cópia assinada pelos pesquisadores responsáveis.

A fase de coletas consistiu em três etapas: a) coleta de sangue periférico, b) aplicação do questionário sobre características de pigmentação e ancestralidade, c) registro fotográfico e mensuração de índice de melanina:

- a) Coletas de sangue: cada indivíduo amostrado doou 10mL de sangue periférico coletado pela equipe de enfermagem do Hemocentro de Ribeirão Preto (unidades USP e Centro) e por técnicos especializados do Departamento de Genética (Laboratório de Genética Bioquímica) da FMRP/USP. O sangue coletado em tubos com EDTA foi armazenado a 4°C para posterior extração do DNA.
- b) Questionário sobre características étnicas e morfológicas: Após a coleta de sangue, foi aplicado pelos membros do laboratório, um questionário onde o doador classificava sua cor de pele, olhos e cabelos, e sua ancestralidade (autoclassificação). Nesse mesmo questionário, dois pesquisadores classificavam as características de pigmentação do doador separadamente (heteroclassificação). A pele foi classificada quanto a cor e facilidade de bronzear (Fitzpatrick 1988) e presença de pintas e sardas. Os olhos foram classificados quanto a cor (azul, verde, mel e castanhos claro e escuro), homogeneidade da cor durante os 20 primeiros anos de vida e presença de heterocromia. Os cabelos foram classificados em relação a cor (loiros claro e escuro, ruivos, castanhos claros e escuros e pretos), homogeneidade da cor durante os primeiros 20 anos de vida e presença de cabelos brancos (0%, 0-25%, 25-50%, 50-75%, 75-100%). A ancestralidade dos avós

maternos e paternos foi classificada em europeia, ameríndia, africana, asiática ou miscigenada (brasileira). Os indivíduos ainda classificaram seus pais e a si mesmos segundo a classificação do IBGE em brancos, pretos, pardos, amarelos ou indígenas.

- c) Registro fotográfico e mensuração de índice de melanina (Espectrofotômetro CM-2500d, Konica Minolta): De cada doador foram tiradas fotos de áreas de alta e baixa exposição solar (testa e interior do antebraço, respectivamente), raiz dos cabelos e olhos, fazendo uso de uma câmera fotográfica SX510HS (Canon U.S.A., Inc.), ajustada com os seguintes parâmetros: abertura $f = 3,4$, *shutterspeed* = 1/10 e ISO = 100. Medidas espectrofotométricas que levam ao cálculo do índice de melanina foram obtidas em triplicata a partir de áreas de alta e baixa exposição solar (testa e interior do antebraço, respectivamente), fazendo uso do espectrofotômetro CM-2500d e do software SpectraMagic NX (Konica Minolta, Inc.). As fotos e medidas foram feitas, sempre que possível, no mesmo local para manter as mesmas condições de iluminação durante a etapa de coleta.

Os questionários, fotos e medidas espectrofotométricas foram armazenados para posterior confronto com os dados obtidos no sequenciamento das amostras.

5.2. Procedimentos Laboratoriais (Anexo C)

Durante o desenvolvimento do projeto, iniciado em 2014 pelo grupo de pesquisa do Laboratório de Pesquisas Forenses e Genômicas, foi elaborado um protocolo contendo os procedimentos laboratoriais seguidos pelo grupo (Anexo C). Aqui será apresentada uma descrição sucinta desses procedimentos.

5.2.1. Extração (Anexo C - Item 1)

O DNA foi extraído por *Salting-out* (Miller et al. 1988), ressuspendido em 200µL de água e armazenado a -20°C. Durante essa etapa não foi possível extrair o DNA de 5 amostras

(PIG0043, PIG0065, PIG0148, PIG0212 e PIG0256). Restando um total de 341 amostras para análise.

5.2.2. Quantificação (Anexo C - Item 2)

O DNA extraído foi diluído em água (1:10) e quantificado por espectrofotometria (*NanoDrop 2000 UV-Vis Spectrophotometer*, Thermo Fisher Scientific Inc.) para avaliar a contaminação por proteínas, e fluorimetria (*Qubit® 2.0 Fluorometer*, Thermo Fisher Scientific Inc.) a fim de preparar alíquotas de 5ng/μL para a etapa de preparação de bibliotecas. Além disso, foi feito eletroforese em gel de agarose das amostras, para identificar eventual degradação do DNA extraído.

5.2.3. Preparação de Bibliotecas (Anexo C - Itens 3 a 5)

As bibliotecas das amostras foram preparadas utilizando o kit *HaloPlex Target Enrichment System* (Agilent Technologies), que consiste em sondas preparadas a partir de um ensaio *in silico* (*SureDesign*, Agilent Technologies) utilizadas seguindo o protocolo da Agilent Technologies para sequenciamento Illumina (*HaloPlex Target Enrichment System For Illumina Sequencing*, Agilent Technologies, 2013) com algumas alterações, descritas detalhadamente no Anexo C. Para o gene *MITF*, as sondas cobrem integralmente a região promotora, 3'UTR e os éxons, além de abrangerem as porções intrônicas que flanqueiam esses éxons.

De maneira resumida, a preparação das bibliotecas consiste em seis etapas:

- 1) Digestão das amostras por enzimas de restrição.
- 2) Hibridização dos fragmentos alvo com sondas.
- 3) Captura dos fragmentos hibridizados.
- 4) Amplificação dos fragmentos alvo por PCR.
- 5) Purificação das bibliotecas.

Os procedimentos de quantificação e avaliação da qualidade das bibliotecas para montagem do pool de até 96 bibliotecas para sequenciamento foram feitos por eletroforese capilar (*Bioanalyzer 2100*, Agilent Technologies), fluorimetria (*Qubit® 2.0 Fluorometer*,

Thermo Fisher Scientific Inc.) e PCR em Tempo Real (*ViiA™ 7 Real-Time PCR System, Applied Biosystems™*). Após a quantificação as bibliotecas foram armazenadas a -20°C.

5.2.4. Sequenciamento (Anexo C - Item 6)

O pool de bibliotecas para o sequenciamento foi preparado de maneira a obter 600µL de uma solução 16pM em que todas as bibliotecas estão presentes na mesma proporção. Essa solução é então inserida no cartucho *MiSeq Reagent Kit V3 (600 cycle)* para sequenciamento no *MiSeq Personal Sequencer (Illumina Inc.)*. Esse preparo seguiu o Guia de Desnaturação e Diluição de Bibliotecas do Sistema *MiSeq* (http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-01.pdf). Nessa etapa uma das 341 amostras (PIG0240) não foi sequenciada, restando 340 amostras.

5.3. Processamento de Dados do Sequenciamento (Anexo D)

Durante o desenvolvimento do projeto, iniciado em 2014 pelo grupo de pesquisa do Laboratório de Pesquisas Forenses e Genômicas, foram elaborados protocolos contendo o fluxograma seguido pelo grupo para o tratamento e análise dos dados de sequenciamento (Anexo D). Aqui será apresentada uma descrição breve desses passos.

5.3.1. Trimagem e Obtenção de Genótipos (Anexo D)

Os arquivos FastQ obtidos no sequenciamento foram trimados utilizando a ferramenta cutadapt (Martin 2011) para retirar os adaptadores utilizados no sequenciamento e sequências de baixa qualidade. As sequências trimadas foram mapeadas utilizando o pacote BWA MEM (Li e Durbin 2009).

Após o mapeamento, as variantes das regiões promotora, UTRs e codificante do gene *MITF* foram identificadas utilizando o pacote GATK (Van der Auwera et al. 2013; DePristo et al. 2011; McKenna et al. 2010). Foram utilizadas tanto a ferramenta *HaplotypeCaller* quanto a *UnifiedGenotyper*. As duas ferramentas possuem modelos e configurações bastante distintos e, por isso, podem resultar em arquivos VCF com variantes diferentes. Os arquivos VCF gerados pelas duas ferramentas foram, então, processados utilizando o *software VCFx* (Castelli et al. 2015), uma ferramenta que interroga variantes em função de critérios pré-estabelecidos de cobertura que corresponderiam a genótipos homozigotos, heterozigotos ou interrogados. Essa ferramenta retorna um arquivo VCF onde os alelos interrogados são substituídos por um ponto. Os dois arquivos gerados foram analisados a fim de determinar quais variantes foram identificadas de maneira correta pelos algoritmos. As variantes identificadas somente por um dos algoritmos foram analisadas manualmente, utilizando o *Integrative Genomics Viewer* 2.3.75 (Robinson et al. 2011; Thorvaldsdottir et al. 2013) a fim de determinar o genótipo correto. Com base nessas análises foi criado um arquivo vcf composto por informações obtidas a partir das duas ferramentas (*HaplotypeCaller* + *VCFx* e *UnifiedGenotyper* + *VCFx*).

Em seguida foi utilizada a ferramenta *ReadBackedPhasing* (GATK) para obter a fase de variantes que se encontram em uma mesma *read*. O arquivo VCF agora possui informações que permitem a criação de um arquivo que indica as fases dessas variantes e auxilia na inferência da fase do restante dos pontos.

5.4. Obtenção dos haplótipos (Anexo E)

O *software PHASE* (Stephens et al. 2001; Stephens e Donnelly 2003) foi utilizado para inferir computacionalmente fases desconhecidas dos haplótipos das amostras sequenciadas. O Protocolo de análise estatística de dados de sequenciamento de nova geração (Anexo E) descreve duas maneiras de inferir haplótipos. A inferência dos haplótipos para o gene *MITF* foi realizada utilizando a abordagem de Faseamento Fragmentado (Anexo E, item 2.2). Nessa abordagem a inferência é feita separadamente para cada região do gene. Para isso, as variantes do gene *MITF* foram agrupadas em 3 regiões: promotora, com 18 sítios de variação; região 5'UTR e codificante, com 38 sítios de variação; região 3'UTR, com 17 sítios de variação. Como descrito no Anexo E, os haplótipos inferidos com probabilidade maior que 80% seriam aceitos como corretos.

Os haplótipos inferidos nas três regiões foram nomeados com o auxílio de rede de haplótipos (*networks*) criadas utilizando o software PopART (disponível em <http://popart.otago.ac.nz>). As *networks* representam graficamente as frequências dos haplótipos e a relação molecular entre eles. Desse modo, os haplótipos de frequências maiores foram considerados como os ancestrais, e os restantes como derivados desse haplótipo. Para a nomenclatura dos haplótipos, cada mutação, ou ponto de variação, entre o haplótipo original e seus derivados foi representada por números e letras intercalados, sendo que cada número e letra representa uma variação. Os haplótipos foram nomeados utilizando os prefixos “Promo” para a promotora, “5’CD” para a 5’UTR e codificante, e “3UTR” para haplótipos na região 3’UTR.

A fim de reconstruir os haplótipos completos do gene *MITF*, os haplótipos agora nomeados foram considerados alelos dentro de cada uma das três regiões (*locus*). Portanto, o arquivo de *input* utilizado no faseamento do gene completo foi composto por três *loci* polimórficos multialélicos: o primeiro *locus* (região promotora) possui 17 alelos (haplótipos), o segundo *locus* (5’UTR + CDS) possui 51 alelos (haplótipos), e o terceiro *locus* (região 3’UTR) possui 18 alelos (haplótipos). Devido a limitação do software PHASE, somente 49 haplótipos da região 5’UTR + CDS puderam ser utilizados no faseamento final, sendo que os dois haplótipos não utilizados só aparecem uma única vez (*singletons*) e pertencem a indivíduos de características fenotípicas mais comuns (olhos e cabelos castanhos e pele média). Haplótipos que não foram inferidos com probabilidade mínima de 80% foram representados por -1, caracterizando um alelo *missing*. Para que o programa não trate os *locus* como STRs (*stepwise mutation model*) foi utilizado o argumento -d1 na linha de comando, como descrito no Anexo E. A inferência dos haplótipos feita pelo programa foi aceita como correta quando a probabilidade foi de pelo menos 80%. Após este processo de faseamento final, os alelos *missing* inferidos foram conferidos manualmente para verificar se a inferência é plausível quando confrontada com os sítios de variação genotipados com sucesso naquele *locus*. Após este processo, os indivíduos que não tiveram seus haplótipos completos definidos com probabilidade mínima de 80% foram descartados das análises seguintes.

5.5. Análises Estatísticas (Anexo E, Itens 3 a 5 e Anexo F)

As amostras foram analisadas a fim de avaliar a diversidade genética do gene *MITF* em suas regiões promotora, UTRs e CDS e encontrar associações significativas entre as características de pigmentação e os genótipos e haplótipos encontrados.

As heterozigoses observada e esperada, aderência ao equilíbrio de Hardy-Weinberg e diversidade molecular foram analisados utilizando Arlequin 3.5.2.2 (Excoffier e Lischer 2010). Essas análises foram feitas para cada região do gene *MITF*, com as amostras estratificadas de acordo com os diferentes fenótipos e globalmente. O desequilíbrio de ligação foi estudado para as regiões do gene *MITF*, assim como para a extensão total do gene, para os diferentes grupos fenotípicos, com o auxílio do software Haploview (Barrett et al. 2005).

Posteriormente, os indivíduos foram estratificados conforme suas características fenotípicas a fim de encontrar associações significativas ($p < 0,05$) e fortes ($OR > 10$) entre alelos das variantes, genótipos e haplótipos e essas características. Para isso foi utilizada uma planilha desenvolvida em nosso laboratório. As associações encontradas foram estudadas com o intuito de entender se alterariam a função ou regulação desse gene. Para isso foram utilizados Haploreg v4.1 (Ward e Kellis 2012) para identificar possíveis alterações funcionais nas regiões regulatórias e SNVrap (Li e Wang 2015) e SnpEff (Cingolani et al. 2012) para identificar possíveis efeitos causados por variantes presentes em regiões codificantes. Haploreg é uma ferramenta online que reúne anotações em regiões não codificantes do genoma (disponível em <http://archive.broadinstitute.org/mammals/haploreg/haploreg.php>). SNVrap é uma ferramenta online de anotação gênica (disponível em <http://jjwanglab.org/snvrap/>) enquanto que SnpEff é um conjunto ferramentas de anotação gênica e predição de efeito de variantes.

A fim de encontrar possíveis miRNAs que possam regular a expressão do gene *MITF* a partir da interação com os diferentes haplótipos encontrados para a região 3'UTR, foi utilizado o software mirhunt (Porto et al. 2015) (Anexo F). Esse software procura por interações entre a sequência alvo e miRNAs utilizando outras 3 ferramentas: RNAhybrid (Rehmsmeier et al. 2004), miRanda (John et al. 2004) e IntaRNA (Busch et al. 2008). As interações encontradas são graduadas segundo sua especificidade (valor de S), estabilidade (valor de E) e uma nota combinada dos dois parâmetros (valor de ES). Essa procura foi realizada para cada um dos 18 haplótipos encontrados para a região 3'UTR e os valores de S, E e ES foram comparados a fim de encontrar os miRNAs que interagem com a região 3'UTR do gene *MITF* de forma mais específica e estável, além de identificar quais desses miRNAs têm sua atividade mais afetada pelas variantes presentes nessa região.

6. Resultados

6. Resultados

6.1. Amostra populacional

A amostragem abrangeu 167 mulheres e 173 homens brasileiros, de idade entre 18 e 72 anos, de diferentes ancestralidades. Quanto às características de pigmentação, foram amostrados 144 indivíduos de pele clara (I e II na escala Fitzpatrick), dos quais 30 possuem sardas, 140 indivíduos de pele moderadamente pigmentada (III e IV na escala Fitzpatrick), dos quais 9 possuem sardas, e 56 indivíduos de pele escura (V e VI na escala Fitzpatrick), dos quais somente 1 apresenta sardas. Quanto à pigmentação da íris, 37 dos 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. Em relação aos cabelos, 11 indivíduos apresentam cabelos ruivos, 13 apresentam cabelos loiro claro e 36 loiro escuro, 61 indivíduos apresentam cabelos castanho claro, 125 possuem cabelos castanho escuro e 86 indivíduos possuem cabelos pretos. Não foi possível determinar a pigmentação dos cabelos de oito indivíduos. Quanto à classificação do IBGE, 221 indivíduos foram classificados pelo grupo de pesquisa como brancos, 64 como pardos, 44 como pretos e 11 como amarelos, enquanto que 218 doadores se auto classificaram como brancos, 67 como pardos, 42 como pretos, 12 como amarelos e 1 como indígena. Os gráficos na Figura 4 mostram a distribuição das características descritas acima na população amostrada.

Dentre os 340 indivíduos sequenciados, há alguns indivíduos aparentados, que serviram como controle de qualidade de todo o processo de *variant calling* e inferência de haplótipos: os irmãos PIG0136, PIG0137 e PIG0138, e PIG0289 e PIG0291; as mães PIG0045 e PIG0329 e seus respectivos filhos PIG0046 e PIG0330; e o trio PIG0310, PIG0311 e PIG0312. Desses, as amostras PIG0046, PIG0136, PIG0138, PIG0291, PIG0311, PIG0312 e PIG0329 foram removidas das análises populacionais, restando 333 amostras. A escolha das amostras a serem removidas foi feita baseado na qualidade dos *callings* (quantidade de *missings*) e características fenotípicas.

Figura 4. Proporções de indivíduos com diferentes fenótipos de pigmentação de pele (a), classificação do IBGE (b), cor de cabelo (c) e cor de olho (d).

6.2. Sequenciamento e tratamento computacional

Foram realizados cinco ensaios nos quais cada uma das 340 bibliotecas foi sequenciada pelo menos uma vez. Amostras que apresentaram cobertura baixa quando sequenciadas pela primeira vez, foram re-sequenciadas em ensaios posteriores. A Tabela 1 mostra a densidade e quantidade de *reads* aproveitadas em cada sequenciamento. Segundo texto “*Optimizing Cluster Density on Illumina Sequencing Systems*” (Illumina, Inc., disponível em <http://www.illumina.com/company/news-center/feature-articles.html>), uma clusterização entre 1200 e 1400K/mm³ seria ideal para o ensaio *MiSeq v3*, equilibrando a qualidade e quantidade dos dados gerados. Clusterização mais alta gera uma quantidade maior de *reads*, aumentando a cobertura, porém pode comprometer a qualidade dessas *reads*, uma vez que a quantidade excessiva de *reads* prejudica a leitura das bases incorporadas, resultando em valores de PHRED mais baixos. Apesar das densidades altas obtidas em cada sequenciamento, o aproveitamento foi em torno de 85%, somando pelo menos 1.500 clusters de boa qualidade por milímetro quadrado, com exceção do terceiro sequenciamento que, apesar da excelente qualidade (96,7% de clusters aproveitados), gerou uma quantidade menor de dados.

Tabela 1. Produção e aproveitamento de *clusters* de *amplicons* na *flow cell* (K) e *reads* em cada um dos cinco sequenciamentos. Dados obtidos pelo *BaseSpace* (*Illumina, Inc*).

	Densidade (aproveitamento)	Nº de <i>Reads</i>
1º Sequenciamento (<i>n</i> = 48)	2.054 K/mm ² (78,8%)	48,5M <i>reads</i>
2º Sequenciamento (<i>n</i> = 96)	2.004 K/mm ² (82,0%)	47,5M <i>reads</i>
3º Sequenciamento (<i>n</i> = 87)	854 K/mm ² (96,7%)	21,3M <i>reads</i>
4º Sequenciamento (<i>n</i> = 80)	1.925 K/mm ² (84,1%)	46,3M <i>reads</i>
5º Sequenciamento (<i>n</i> = 61)	1.811 K/mm ² (85,7%)	43,7M <i>reads</i>

É comum no ensaio de sequenciamento *MiSeq* da *Illumina* que haja uma queda na qualidade das bases incorporadas nos últimos ciclos, principalmente em *amplicons* menores que o total de ciclos de sequenciamento (301) (Yang et al. 2013; Erlich et al. 2008). Porém, as *reads* foram processadas a fim de remover os adaptadores presentes em suas extremidades. Além disso, o software cutadapt remove integralmente as *reads* que não foram trimadas (que não possuíam identificação, como adaptadores e índices). Como mostrado na Tabela 2, a porção restante das *reads* possui uma qualidade média mais alta que das bases das *reads* nos dados brutos, e a quantidade de *reads* descartadas corresponde a menos de 5% do total.

Tabela 2. Distribuição de *reads* de acordo com a qualidade média das bases que as compõem, pré e pós trimagem dos adaptadores *paired-end* da *Illumina*. Os dados representando cada sequenciamento são, respectivamente, das amostras PIG0001, PIG0100, PIG0261, PIG0307 e PIG0341, as quais foram tomadas aleatoriamente. Dados obtidos com FastQC (Andrews 2010).

		PHRED score		
		< 20	20 < x < 30	≥ 30
1º sequenciamento <i>n</i> =48	pré trimagem	26.003 (3,4%)	407.581 (54,7%)	311.031 (41,7%)
	pós trimagem	5.600 (0,7%)	49.156 (6,8%)	658.521 (92,3%)
2º sequenciamento <i>n</i> =96	pré trimagem	12.491 (3,9%)	187.314 (58,6%)	119.417 (37,4%)
	pós trimagem	6.656 (2,1%)	16.611 (5,3%)	288.331 (92,5%)
3º sequenciamento <i>n</i> =87	pré trimagem	1.849 (0,9%)	64.931 (31,9%)	136.377 (67,1%)
	pós trimagem	1.483 (0,7%)	8.047 (4,0%)	190.555 (95,2%)
4º sequenciamento <i>n</i> =80	pré trimagem	2.439 (0,7%)	163.240 (52,1%)	147.425 (47,0%)
	pós trimagem	1.000 (0,3%)	16.877 (5,5%)	286.369 (94,1%)
5º sequenciamento <i>n</i> =61	pré trimagem	4.628 (1,1%)	154.957 (38,1%)	246.567 (60,7%)
	pós trimagem	1.723 (0,4%)	27.111 (6,7%)	372.679 (92,8%)

Após a trimagem, as *reads* foram mapeadas a um genoma humano de referência (hg19). Como mostrado na Tabela 3, os cinco ensaios tiveram em torno de 89% das *reads* mapeadas com qualidade mínima Q30 (probabilidade de erro de 1 em 1000).

Tabela 3. Qualidade de mapeamento ao genoma de referência. Os dados representando cada sequenciamento (lote) são, respectivamente, das amostras PIG0001, PIG0100, PIG0261, PIG0307 e PIG0341, as quais foram tomadas aleatoriamente. Dados obtidos com SAMStat (Lassmann et al. 2011).

PHRED score	1° lote n=48	2° lote n=96	3° lote n=87	4° lote n=80	5° lote n=61
>=30	1.270.414 (88,8%)	546.220 (87,5%)	356.713 (88,0%)	569.837 (90,7%)	745.211 (91,9%)
20<x<30	5.721 (0,4%)	2.318 (0,4%)	1.498 (0,4%)	2.234 (0,4%)	3.615 (0,4%)
10<x<20	3.512 (0,2%)	1.441 (0,2%)	1.457 (0,4%)	955 (0,2%)	1.365 (0,2%)
3<x<10	3.167 (0,2%)	1.960 (0,3%)	1.775 (0,4%)	1.062 (0,2%)	1.437,2%)
>10	434 (0%)	197 (0%)	358 (0,1%)	296 (0%)	497 (0,1%)
Não mapeadas	147.241 (10,3%)	72.271 (11,6%)	43.615 (10,8%)	53.561 (8,5%)	58.992 (7,3%)
Total	1.430.489 (100%)	624.407 (100%)	405.416 (100%)	627.945 (100%)	811.11700%)

6.3. Variant Calling

Os algoritmos *HaplotypeCaller* e *UnifiedGenotyper* (GATK) encontraram 131 e 133 pontos de variação nas regiões cobertas do gene *MITF*, respectivamente, sendo que 121 desses pontos foram encontrados por ambos. Os *callings* feitos pelos dois algoritmos foram analisados com o auxílio do software VCFx v0.10.1, que interroga alelos de acordo com algumas regras descritas no Anexo D, item 1.3.2 que levam em consideração a proporção de alelos referência e alternativo, além da cobertura total para determinado indivíduo em determinada posição. Esse processo garante que, em regiões de baixa cobertura, genótipos homozigotos sejam determinados somente quando realmente não houver variação e evita o *calling* de genótipos heterozigotos em que o alelo alternativo tem uma cobertura muito baixa e pode ser proveniente de algum erro de mapeamento ou *base calling* (Castelli et al. 2015). Os alelos interrogados foram inferidos na etapa de faseamento (PHASE).

Após essa análise foi criado um VCF final composto por 119 das variantes encontradas pelos dois algoritmos, 10 *indels* encontrados pelo *HaplotypeCaller* e 4 outras variantes encontradas somente pelo *UnifiedGenotyper*, somando 133 variantes (Apêndice A), das quais 58 são *singletons* e 10 são *indels*. As 133 variantes se distribuíram da seguinte forma (Figura 5): 30 (12 *singletons*) estão presentes na região promotora (chr3:69785586-69788585), 3 (2 *singletons*) na região 5'UTR (chr3:69788586-69788748), 68 (30 *singletons*) na região codificante (chr3:69788749-70014399), sendo 8 variantes em regiões codificantes de proteína e 60 variantes em regiões não codificantes, e 32 variantes (14 *singletons*) na região 3'UTR (chr3:70014400-70017488).

Figura 5. Distribuição dos 133 pontos de variação em toda a extensão do gene *MITF*. A região codificadora (CD) corresponde às regiões exônicas adicionadas de fragmentos flanqueadores de regiões intrônicas cobertas pelo ensaio Haloplex.

6.4. Frequências alélicas, genotípicas, haplotípicas e heterozigose

As estatísticas populacionais foram calculadas para o grupo de indivíduos não aparentados amostrados aleatoriamente. Nessa amostragem ($n = 198$), somente 33 das variantes são polimórficas (alelo menos frequente com frequência mínima de 1%), 67 têm seus alelos menos frequentes com frequência entre 0 e 1% e 33 posições não são variáveis. A heterozigose média observada foi de $0,0704 \pm 0,1423$ e a heterozigose média esperada foi de $0,0676 \pm 0,1329$. As variantes rs1461187921 e rs2217456 desviam significativamente do equilíbrio de Hardy-Weinberg ($p = 0,0039 \pm 0,0001$ e $p = 0,0112 \pm 0,0001$), porém somente a variante rs2217456 está associada aos fenótipos estudados, como será visto posteriormente.

As Tabelas 4, 5 e 6 mostram as frequências alélicas e genotípicas das variantes que foram associadas a algum fenótipo de pigmentação, tanto para a amostragem aleatória ($n = 198$) quanto para o total de indivíduos não aparentados ($n = 333$). Nessa amostragem de 333 indivíduos não aparentados, 60 das 133 variantes são *singletons*.

Para a reconstrução dos haplótipos foram considerados 73 pontos de variação em todo o gene *MITF* e região promotora, sendo 18 posições na região promotora, uma posição na região 5'UTR, 37 posições na região codificante e 17 posições dentro da região 3'UTR. Foram encontrados 17 haplótipos na região promotora, 50 haplótipos que abrangem a região 5'UTR e codificante, 18 haplótipos na região 3'UTR e 132 haplótipos foram encontrados considerando a extensão total. A distribuição haplotípica nas regiões individualizadas do gene, assim como na extensão total, estão de acordo com o esperado pelo modelo do equilíbrio de Hardy-Weinberg. Após a inferência dos haplótipos das regiões promotora, 5'UTR e CD e 3'UTR,

devido a remoção dos indivíduos que tiveram seus haplótipos inferidos com probabilidade menor que 80%, cinco pontos de variação se mostraram não polimórficos, um na região promotora, dois na região CD e dois na região 3'UTR. As estatísticas referentes aos haplótipos foram calculadas para cada região do gene *MITF* e para a extensão total.

Tabela 4. Frequências alélicas e genotípicas, heterozigose observada (H_o) e esperada (H_e), e aderência ao Equilíbrio de Hardy-Weinberg ($pEHW$) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de p significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, intron, éxon e 3'UTR, respectivamente. (continua)

Ponto de variação		Pele clara		Pele média		Pele escura		Presença de sardas		Presença de sardas e pintas		Ausência de sardas e pintas		Total	
		2n=142	2n=286	2n=204	2n=268	2n=50	2n=112	2n=32	2n=80	2n=296	2n=502	2n=100	2n=164	2n=396	2n=666
rs74845300 ¹ (69786714)	A/A	1,0000	1,0000	1,0000	0,9925	0,9600	0,9107	1,0000	1,0000	0,9932	0,9880	1,0000	0,9634	0,9949	0,9820
	A/G	0,0000	0,0000	0,0000	0,0075	0,0400	0,0893	0,0000	0,0000	0,0068	0,0120	0,0000	0,0366	0,0051	0,0180
	G/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	1,0000	1,0000	0,9963	0,9800	0,9554	1,0000	1,0000	0,9966	0,9940	1,0000	0,9817	0,9975	0,9910
	G	0,0000	0,0000	0,0000	0,0037	0,0200	0,0446	0,0000	0,0000	0,0034	0,0060	0,0000	0,0183	0,0025	0,0090
	H_o	-	-	-	0,0075	0,0400	0,0893	-	-	0,0068	0,0120	-	0,0366	0,0051	0,0180
	H_e	-	-	-	0,0075	0,0400	0,0861	-	-	0,0068	0,0119	-	0,0361	0,0051	0,0179
	$pEHW$	-	-	-	1,0000	1,0000	1,0000	-	-	1,0000	1,0000	-	1,0000	1,0000	1,0000
rs80232752 ¹ (69787573)	A/A	1,0000	1,0000	1,0000	0,9925	0,9600	0,9107	1,0000	1,0000	0,9932	0,9880	1,0000	0,9634	0,9949	0,9820
	A/T	0,0000	0,0000	0,0000	0,0075	0,0400	0,0893	0,0000	0,0000	0,0068	0,0120	0,0000	0,0366	0,0051	0,0180
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	1,0000	1,0000	0,9963	0,9800	0,9554	1,0000	1,0000	0,9966	0,9940	1,0000	0,9817	0,9975	0,9910
	T	0,0000	0,0000	0,0000	0,0037	0,0200	0,0446	0,0000	0,0000	0,0034	0,0060	0,0000	0,0183	0,0025	0,0090
	H_o	-	-	-	0,0075	0,0400	0,0893	-	-	0,0068	0,0120	-	0,0366	0,0051	0,0180
	H_e	-	-	-	0,0075	0,0400	0,0861	-	-	0,0068	0,0119	-	0,0361	0,0051	0,0179
	$pEHW$	-	-	-	1,0000	1,0000	1,0000	-	-	1,0000	1,0000	-	1,0000	1,0000	1,0000
rs12489786 ¹ (69788068)	T/T	0,9859	0,9930	0,9314	0,9328	1,0000	0,9286	1,0000	0,9750	0,9662	0,9641	0,9400	0,9390	0,9596	0,9580
	T/C	0,0141	0,0070	0,0588	0,0597	0,0000	0,0714	0,0000	0,0250	0,0270	0,0319	0,0600	0,0610	0,0354	0,0390
	C/C	0,0000	0,0000	0,0098	0,0075	0,0000	0,0000	0,0000	0,0000	0,0068	0,0040	0,0000	0,0000	0,0051	0,0030
	T	0,9930	0,9965	0,9608	0,9627	1,0000	0,9643	1,0000	0,9875	0,9797	0,9801	0,9700	0,9695	0,9773	0,9775
	C	0,0070	0,0035	0,0392	0,0373	0,0000	0,0357	0,0000	0,0125	0,0203	0,0199	0,0300	0,0305	0,0227	0,0225
	H_o	0,0141	0,0070	0,0588	0,0597	-	0,0714	-	0,0250	0,0270	0,0319	0,0600	0,0610	0,0354	0,0390
	H_e	0,0141	0,0070	0,0757	0,0721	-	0,0695	-	0,0250	0,0399	0,0391	0,0588	0,0595	0,0445	0,0441
	$pEHW$	1,0000	1,0000	0,1324	0,1588	-	1,0000	-	1,0000	0,0502	0,0878	1,0000	1,0000	0,0892	0,1480

Tabela 4. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, intron, éxon e 3'UTR, respectivamente.
(continuação)

Ponto de variação		Pele clara		Pele média		Pele escura		Presença de sardas		Presença de sardas e pintas		Ausência de sardas e pintas		Total	
		2n=142	2n=286	2n=204	2n=268	2n=50	2n=112	2n=32	2n=80	2n=296	2n=502	2n=100	2n=164	2n=396	2n=666
rs531268453 ¹ (69788550)	C/C	1,0000	0,9930	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9960	1,0000	1,0000	1,0000	0,9970
	C/T	0,0000	0,0070	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0040	0,0000	0,0000	0,0000	0,0030
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	0,9965	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9980	1,0000	1,0000	1,0000	0,9985
	T	0,0000	0,0035	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0020	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	<i>He</i>	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	<i>pEHW</i>	-	1,0000	-	-	-	-	-	-	-	1,0000	-	-	-	1,0000
rs114935806 ¹ (69788560)	C/C	1,0000	1,0000	1,0000	1,0000	1,0000	0,9643	1,0000	1,0000	1,0000	1,0000	1,0000	0,9756	1,0000	0,9940
	C/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0357	0,0000	0,0000	0,0000	0,0000	0,0000	0,0244	0,0000	0,0060
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	1,0000	1,0000	1,0000	1,0000	0,9821	1,0000	1,0000	1,0000	1,0000	1,0000	0,9878	1,0000	0,9970
	T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0179	0,0000	0,0000	0,0000	0,0000	0,0000	0,1220	0,0000	0,0030
	<i>Ho</i>	-	-	-	-	-	0,0357	-	-	-	-	-	0,0244	-	0,0060
	<i>He</i>	-	-	-	-	-	0,0354	-	-	-	-	-	0,0242	-	0,0060
	<i>pEHW</i>	-	-	-	-	-	1,0000	-	-	-	-	-	1,0000	-	1,0000
rs561386441 ¹ (69788574)	C/C	1,0000	1,0000	0,9902	0,9925	1,0000	0,9464	1,0000	1,0000	0,9932	0,9920	1,0000	0,9756	0,9949	0,9880
	C/G	0,0000	0,0000	0,0098	0,0075	0,0000	0,0536	0,0000	0,0000	0,0068	0,0080	0,0000	0,0244	0,0051	0,0120
	G/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	1,0000	0,9951	0,9963	1,0000	0,9732	1,0000	1,0000	0,9966	0,9960	1,0000	0,9878	0,9975	0,9940
	G	0,0000	0,0000	0,0049	0,0037	0,0000	0,0268	0,0000	0,0000	0,0034	0,0040	0,0000	0,0122	0,0025	0,0060
	<i>Ho</i>	-	-	0,0098	0,0075	-	0,0536	-	-	0,0068	0,0080	-	0,0244	0,0051	0,0120
	<i>He</i>	-	-	0,0098	0,0075	-	0,0526	-	-	0,0068	0,0080	-	0,0242	0,0051	0,0120
	<i>pEHW</i>	-	-	1,0000	1,0000	-	1,0000	-	-	1,0000	1,0000	-	1,0000	1,0000	1,0000

Tabela 4. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, intron, éxon e 3'UTR, respectivamente. (continuação)

Ponto de variação		Pele clara		Pele média		Pele escura		Presença de sardas		Presença de sardas e pintas		Ausência de sardas e pintas		Total	
		2n=142	2n=286	2n=204	2n=268	2n=50	2n=112	2n=32	2n=80	2n=296	2n=502	2n=100	2n=164	2n=396	2n=666
rs113085085 ² (69801287)	A/A	1,0000	0,9930	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9960	1,0000	1,0000	1,0000	0,9970
	A/C	0,0000	0,0070	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0040	0,0000	0,0000	0,0000	0,0030
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	0,9965	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9980	1,0000	1,0000	1,0000	0,9985
	C	0,0000	0,0035	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0020	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	<i>He</i>	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	<i>pEHW</i>	-	1,0000	-	-	-	-	-	-	-	1,0000	-	-	-	1,0000
rs13322505 ² (69801975)	T/T	1,0000	1,0000	1,0000	0,9925	0,8400	0,9107	1,0000	1,0000	0,9932	0,9920	0,9400	0,9512	0,9798	0,9819
	T/C	0,0000	0,0000	0,0000	0,0075	0,1600	0,0893	0,0000	0,0000	0,0068	0,0080	0,0600	0,0488	0,0202	0,0181
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	T	1,0000	1,0000	1,0000	0,9925	0,9200	0,9554	1,0000	1,0000	0,9966	0,9940	0,9700	0,9756	0,9899	0,9895
	C	0,0000	0,0000	0,0000	0,0075	0,0800	0,0446	0,0000	0,0000	0,0034	0,0060	0,0300	0,0244	0,0101	0,0105
	<i>Ho</i>	-	-	-	0,0075	0,1600	0,0893	-	-	0,0068	0,0080	0,0600	0,4878	0,0202	0,0181
	<i>He</i>	-	-	-	0,0075	0,1502	0,0861	-	-	0,0068	0,0080	0,0588	0,4788	0,0201	0,0179
	<i>pEHW</i>	-	-	-	1,0000	1,0000	1,0000	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
rs76992570 ² (69807750)	G/G	1,0000	1,0000	1,0000	1,0000	1,0000	0,9630	1,0000	1,0000	1,0000	0,9958	1,0000	0,9872	1,0000	0,9937
	G/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0370	0,0000	0,0000	0,0000	0,0042	0,0000	0,0128	0,0000	0,0063
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	1,0000	1,0000	1,0000	1,0000	1,0000	0,9817	1,0000	1,0000	1,0000	0,9980	1,0000	0,9937	1,0000	0,9969
	A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0184	0,0000	0,0000	0,0000	0,0020	0,0000	0,0063	0,0000	0,0031
	<i>Ho</i>	-	-	-	-	-	0,3704	-	-	-	0,0042	-	0,0128	-	0,0635
	<i>He</i>	-	-	-	-	-	0,3669	-	-	-	0,0042	-	0,0128	-	0,0634
	<i>pEHW</i>	-	-	-	-	-	1,0000	-	-	-	1,0000	-	1,0000	-	1,0000

Tabela 4. Frequências alélicas e genotípicas, heterozigose observada (H_o) e esperada (H_e), e aderência ao Equilíbrio de Hardy-Weinberg ($pEHW$) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de p significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, intron, éxon e 3'UTR, respectivamente. (continuação)

Ponto de variação		Pele clara		Pele média		Pele escura		Presença de sardas		Presença de sardas e pintas		Ausência de sardas e pintas		Total	
		2n=142	2n=286	2n=204	2n=268	2n=50	2n=112	2n=32	2n=80	2n=296	2n=502	2n=100	2n=164	2n=396	2n=666
rs80163893 ² (69812860)	C/C	0,9859	0,9860	1,0000	0,9925	1,0000	1,0000	1,0000	0,9750	0,9932	0,9880	1,0000	1,0000	0,9949	0,9910
	C/T	0,0141	0,0140	0,0000	0,0075	0,0000	0,0000	0,0000	0,0250	0,0068	0,0120	0,0000	0,0000	0,0051	0,0090
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	0,9930	0,9930	1,0000	0,9963	1,0000	1,0000	1,0000	0,9875	0,9966	0,9940	1,0000	1,0000	0,9975	0,9955
	T	0,0070	0,0070	0,0000	0,0037	0,0000	0,0000	0,0000	0,0125	0,0034	0,0060	0,0000	0,0000	0,0025	0,0045
	H_o	0,0141	0,0140	-	0,0075	-	-	-	0,0250	0,0068	0,0120	-	-	0,0051	0,0090
	H_e	0,0141	0,0139	-	0,0075	-	-	-	0,0250	0,0068	0,0119	-	-	0,0051	0,0090
	$pEHW$	1,0000	1,0000	-	1,0000	-	-	-	1,0000	1,0000	1,0000	-	-	1,0000	1,0000
rs72950039 ² (69812903)	C/C	0,9859	0,9930	0,9706	0,9701	0,9200	0,9286	0,9375	0,9750	0,9662	0,9721	0,9800	0,9756	0,9697	0,9730
	C/T	0,0141	0,0070	0,0294	0,0299	0,0800	0,0714	0,0625	0,0250	0,0338	0,0279	0,0200	0,0244	0,0303	0,0270
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	0,9930	0,9965	0,9853	0,9851	0,9600	0,9643	0,9688	0,9875	0,9831	0,9861	0,9900	0,9878	0,9849	0,9865
	T	0,0070	0,0035	0,0147	0,0149	0,0400	0,0357	0,0313	0,0125	0,0169	0,0139	0,0100	0,0122	0,0152	0,0135
	H_o	0,0141	0,0070	0,0294	0,0299	0,0800	0,0714	0,0625	0,0250	0,0338	0,0279	0,0200	0,0244	0,0303	0,0270
	H_e	0,0141	0,0070	0,0291	0,0295	0,0784	0,0695	0,0625	0,0250	0,0333	0,0276	0,0200	0,0242	0,0299	0,0267
	$pEHW$	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
rs2217456 ² (69887578)	A/A	0,1692	0,1591	0,2755	0,3178	0,3600	0,4000	0,1875	0,2105	0,2199	0,2395	0,3404	0,3462	0,2500	0,2658
	A/G	0,4923	0,4621	0,3980	0,4031	0,2000	0,3455	0,6250	0,5000	0,4397	0,4328	0,2979	0,3718	0,4043	0,4177
	G/G	0,3385	0,3788	0,3265	0,2791	0,4400	0,2545	0,1875	0,2895	0,3404	0,3277	0,3617	0,2821	0,3457	0,3165
	A	0,4254	0,3911	0,4747	0,5172	0,4600	0,5766	0,5000	0,4545	0,4386	0,4343	0,5052	0,5438	0,4555	0,4743
	G	0,5746	0,6089	0,5253	0,4828	0,5400	0,4234	0,5000	0,5455	0,5614	0,5487	0,4948	0,4562	0,5445	0,5257
	H_o	0,4923	0,4621	0,3980	0,4031	0,2000	0,3455	0,6250	0,5000	0,4397	0,4328	0,2979	0,3718	0,4043	0,4177
	H_e	0,4895	0,4777	0,5013	0,5012	0,5069	0,4939	0,5161	0,5035	0,4945	0,4972	0,5052	0,5012	0,4967	0,4995
	$pEHW$	1,0000	0,7167	0,0452	0,0344	0,0035	0,0303	0,6166	1,0000	0,2291	0,0506	0,0079	0,0252	0,0112	0,0048

Tabela 4. Frequências alélicas e genotípicas, heterozigose observada (H_o) e esperada (H_e), e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de p significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, intron, éxon e 3'UTR, respectivamente. (continuação)

Ponto de variação		Pele clara		Pele média		Pele escura		Presença de sardas		Presença de sardas e pintas		Ausência de sardas e pintas		Total	
		2n=142	2n=286	2n=204	2n=268	2n=50	2n=112	2n=32	2n=80	2n=296	2n=502	2n=100	2n=164	2n=396	2n=666
rs78124318 ² (69887807)	G/G	0,8592	0,8741	0,9510	0,9328	0,9600	0,9821	0,8750	0,9250	0,9122	0,9084	0,9400	0,9390	0,9192	0,9159
	G/A	0,1408	0,1259	0,0490	0,0672	0,0400	0,0179	0,1250	0,0750	0,0878	0,0916	0,0600	0,0610	0,0808	0,0841
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	0,9296	0,9371	0,9755	0,9664	0,9800	0,9911	0,9375	0,9625	0,9561	0,9542	0,9700	0,9695	0,9596	0,9580
	A	0,0704	0,0629	0,0245	0,0336	0,0200	0,0089	0,0625	0,0375	0,0439	0,0458	0,0300	0,0305	0,0404	0,0420
	H_o	0,1409	0,1259	0,0490	0,0672	0,0400	0,0179	0,1250	0,0750	0,0878	0,0916	0,0600	0,0610	0,0808	0,0841
	H_e	0,1319	0,1184	0,0481	0,0652	0,0400	0,0179	0,1210	0,0731	0,0843	0,0876	0,0588	0,0595	0,0777	0,0807
	p_{EHW}	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
rs112598669 ² (69985718)	G/G	1,0000	1,0000	0,9902	0,9925	0,9600	0,9286	1,0000	1,0000	0,9932	0,9960	0,9800	0,9512	0,9898	0,9849
	G/A	0,0000	0,0000	0,0098	0,0075	0,0400	0,0714	0,0000	0,0000	0,0068	0,0040	0,0200	0,0488	0,0102	0,0151
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	1,0000	1,0000	0,9951	0,9963	0,9800	0,9643	1,0000	1,0000	0,9966	0,9980	0,9900	0,9756	0,9949	0,9925
	A	0,0000	0,0000	0,0049	0,0037	0,0200	0,0357	0,0000	0,0000	0,0034	0,0020	0,0100	0,0244	0,0051	0,0075
	H_o	-	-	0,0098	0,0075	0,0400	0,0714	-	-	0,0068	0,0040	0,0200	0,0488	0,0102	0,0151
	H_e	-	-	0,0098	0,0075	0,0400	0,0695	-	-	0,0068	0,0040	0,0200	0,0479	0,0101	0,0150
	p_{EHW}	-	-	1,0000	1,0000	1,0000	1,0000	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
rs576634776 ³ (70008502)	C/C	1,0000	0,9930	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9960	1,0000	1,0000	1,0000	0,9970
	C/T	0,0000	0,0070	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0040	0,0000	0,0000	0,0000	0,0030
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	0,9965	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9980	1,0000	1,0000	1,0000	0,9985
	T	0,0000	0,0035	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0020	0,0000	0,0000	0,0000	0,0015
	H_o	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	H_e	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	p_{EHW}	-	1,0000	-	-	-	-	-	-	-	1,0000	-	-	-	1,0000

Tabela 4. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, intron, éxon e 3'UTR, respectivamente. (continuação)

Ponto de variação		Pele clara		Pele média		Pele escura		Presença de sardas		Presença de sardas e pintas		Ausência de sardas e pintas		Total	
		2n=142	2n=286	2n=204	2n=268	2n=50	2n=112	2n=32	2n=80	2n=296	2n=502	2n=100	2n=164	2n=396	2n=666
70014320 ³	C/C	1,0000	0,9930	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9960	1,0000	1,0000	1,0000	0,9970
	C/T	0,0000	0,0070	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0040	0,0000	0,0000	0,0000	0,0030
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	0,9965	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9980	1,0000	1,0000	1,0000	0,9985
	T	0,0000	0,0035	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0020	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	<i>He</i>	-	0,0070	-	-	-	-	-	-	-	0,0040	-	-	-	0,0030
	<i>pEHW</i>	-	1,0000	-	-	-	-	-	-	-	1,0000	-	-	-	1,0000
rs36118030 ³ (70014384)	G/G	1,0000	1,0000	1,0000	1,0000	0,9600	0,9464	1,0000	1,0000	0,9932	0,9920	1,0000	0,9878	0,9949	0,9910
	G/A	0,0000	0,0000	0,0000	0,0000	0,0400	0,0536	0,0000	0,0000	0,0068	0,0080	0,0000	0,0122	0,0051	0,0090
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	1,0000	1,0000	1,0000	1,0000	0,9800	0,9732	1,0000	1,0000	0,9966	0,9960	1,0000	0,9939	0,9975	0,9955
	A	0,0000	0,0000	0,0000	0,0000	0,0200	0,0268	0,0000	0,0000	0,0034	0,0040	0,0000	0,0061	0,0025	0,0045
	<i>Ho</i>	-	-	-	-	0,0400	0,0536	-	-	0,0068	0,0080	-	0,0122	0,0051	0,0090
	<i>He</i>	-	-	-	-	0,0400	0,0526	-	-	0,0068	0,0080	-	0,0122	0,0051	0,0090
	<i>pEHW</i>	-	-	-	-	1,0000	1,0000	-	-	1,0000	1,0000	-	1,0000	1,0000	1,0000
rs150314710 ⁴ (70015549)	T/T	0,9577	0,9580	0,9706	0,9701	0,9600	0,9643	1,0000	1,0000	0,9730	0,9761	0,9400	0,9268	0,9646	0,9640
	T/C	0,0423	0,0420	0,0294	0,0299	0,0400	0,0357	0,0000	0,0000	0,0270	0,0239	0,0600	0,0732	0,0354	0,0360
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	T	0,9789	0,9790	0,9853	0,9851	0,9800	0,9982	1,0000	1,0000	0,9865	0,9881	0,9700	0,9634	0,9823	0,9820
	C	0,0211	0,0210	0,0147	0,0149	0,0200	0,0179	0,0000	0,0000	0,0135	0,0120	0,0300	0,0366	0,0177	0,1802
	<i>Ho</i>	0,0423	0,0420	0,0294	0,0299	0,0400	0,0357	-	-	0,0270	0,0239	0,0600	0,0732	0,0354	0,0360
	<i>He</i>	0,0417	0,0412	0,0291	0,0295	0,0400	0,0354	-	-	0,0268	0,0237	0,0588	0,0709	0,0348	0,0354
	<i>pEHW</i>	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabela 4. Frequências alélicas e genotípicas, heterozigose observada (H_o) e esperada (H_e), e aderência ao Equilíbrio de Hardy-Weinberg ($pEHW$) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo características da pele. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de p significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, intron, éxon e 3'UTR, respectivamente. (conclusão)

Ponto de variação		Pele clara		Pele média		Pele escura		Presença de sardas		Presença de sardas e pintas		Ausência de sardas e pintas		Total	
		2n=142	2n=286	2n=204	2n=268	2n=50	2n=112	2n=32	2n=80	2n=296	2n=502	2n=100	2n=164	2n=396	2n=666
rs2131025 ⁴ (70015647)	T/T	0,7746	0,7692	0,8922	0,8881	0,8400	0,9286	0,9375	0,7000	0,8446	0,8287	0,8400	0,8902	0,8434	0,8438
	T/C	0,2254	0,2308	0,0980	0,1045	0,1200	0,0536	0,0625	0,3000	0,1486	0,1673	0,1400	0,0976	0,1465	0,1502
	C/C	0,0000	0,0000	0,0098	0,0075	0,0400	0,0179	0,0000	0,0000	0,0068	0,0040	0,0200	0,0122	0,0101	0,0060
	T	0,8873	0,8846	0,9412	0,9403	0,9000	0,9554	0,9688	0,8500	0,9189	0,9124	0,9100	0,9390	0,9167	0,9189
	C	0,1127	0,1154	0,0588	0,0597	0,1000	0,0446	0,0313	0,1500	0,0811	0,8765	0,0900	0,0610	0,0833	0,0811
	H_o	0,2254	0,2308	0,0980	0,1045	0,1200	0,0536	0,0625	0,3000	0,1487	0,1673	0,1400	0,0976	0,1465	0,1502
	H_e	0,2014	0,2049	0,1113	0,1127	0,1837	0,0861	0,0625	0,2582	0,1495	0,1603	0,1655	0,1152	0,1532	0,1492
	$pEHW$	0,5860	0,2161	0,2915	0,3787	0,1976	0,0890	1,0000	0,5675	1,0000	0,7032	0,3268	0,2527	0,6282	1,0000
rs78240629 ⁴ (70016569)	A/A	1,0000	1,0000	1,0000	1,0000	0,9600	0,9464	1,0000	1,0000	0,9932	0,9920	1,0000	0,9878	0,9949	0,9910
	A/G	0,0000	0,0000	0,0000	0,0000	0,0400	0,0536	0,0000	0,0000	0,0068	0,0080	0,0000	0,0122	0,0051	0,0090
	G/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	1,0000	1,0000	1,0000	0,9800	0,9732	1,0000	1,0000	0,9966	0,9960	1,0000	0,9939	0,9975	0,9955
	G	0,0000	0,0000	0,0000	0,0000	0,0200	0,0268	0,0000	0,0000	0,0034	0,0040	0,0000	0,0061	0,0025	0,0045
	H_o	-	-	-	-	0,0400	0,0536	-	-	0,0068	0,0080	-	0,0122	0,0051	0,0090
	H_e	-	-	-	-	0,0400	0,0526	-	-	0,0068	0,0080	-	0,0122	0,0051	0,0090
	$pEHW$	-	-	-	-	1,0000	1,0000	-	-	1,0000	1,0000	-	1,0000	1,0000	1,0000
rs77962238 ⁴ (70016857)	G/G	0,9577	0,9580	0,9706	0,9701	0,9600	0,9643	1,0000	1,0000	0,9730	0,9761	0,9400	0,9268	0,9646	0,9640
	G/A	0,0423	0,0420	0,0294	0,0299	0,0400	0,0357	0,0000	0,0000	0,0270	0,0239	0,0600	0,0732	0,0354	0,0360
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	0,9789	0,9790	0,9853	0,9851	0,9800	0,9821	1,0000	1,0000	0,9865	0,9881	0,9700	0,9634	0,9823	0,9820
	A	0,0211	0,0210	0,0147	0,0149	0,0200	0,0179	0,0000	0,0000	0,0135	0,0120	0,0300	0,0366	0,0177	0,0180
	H_o	0,0423	0,0420	0,0294	0,0299	0,0400	0,0357	-	-	0,0270	0,0239	0,0600	0,0732	0,0354	0,0360
	H_e	0,0417	0,0412	0,0291	0,0295	0,0400	0,0354	-	-	0,0268	0,0237	0,0588	0,0709	0,0348	0,0354
	$pEHW$	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabela 5. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continua)

Ponto de variação		Olho azul		Olho verde		Olho mel		Olho castanho claro		Olho castanho escuro		Total	
		2n= 26	2n=74	2n= 44	2n=122	2n= 16	2n= 38	2n= 120	2n= 142	2n= 190	2n=290	2n= 396	2n= 666
rs74845300 ¹ (69786714)	A/A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9895	0,9586	0,9949	0,9820
	A/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0105	0,0414	0,0051	0,0180
	G/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9947	0,9793	0,9975	0,9910
	G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0053	0,2069	0,0025	0,0090
	<i>Ho</i>	-	-	-	-	-	-	-	-	0,0105	0,0414	0,0051	0,0180
	<i>He</i>	-	-	-	-	-	-	-	-	0,0105	0,0407	0,0051	0,0179
	<i>pEHW</i>	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs80232752 ¹ (69787573)	A/A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9895	0,9586	0,9949	0,9820
	A/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0105	0,0414	0,0051	0,0180
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9947	0,9793	0,9975	0,9910
	T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0053	0,0207	0,0025	0,0090
	<i>Ho</i>	-	-	-	-	-	-	-	-	0,0105	0,0414	0,0051	0,0180
	<i>He</i>	-	-	-	-	-	-	-	-	0,0105	0,0407	0,0051	0,0179
	<i>pEHW</i>	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs12489786 ¹ (69788068)	T/T	1,0000	1,0000	0,9545	0,9508	1,0000	0,9474	0,9500	0,9577	0,9579	0,9517	0,9596	0,9580
	T/C	0,0000	0,0000	0,0455	0,0492	0,0000	0,0526	0,0500	0,0423	0,0316	0,0414	0,0354	0,0390
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0105	0,0069	0,0051	0,0030
	T	1,0000	1,0000	0,9773	0,9754	1,0000	0,9737	0,9750	0,9789	0,9737	0,9724	0,9773	0,9775
	C	0,0000	0,0000	0,0227	0,0246	0,0000	0,0263	0,0250	0,0211	0,0263	0,0276	0,0227	0,0225
	<i>Ho</i>	-	-	0,0455	0,0492	-	0,0526	0,0500	0,0423	0,0316	0,0414	0,0354	0,0390
	<i>He</i>	-	-	0,0455	0,0484	-	0,0526	0,0492	0,0417	0,0515	0,0538	0,0445	0,0441
	<i>pEHW</i>	-	-	1,0000	1,0000	-	1,0000	1,0000	1,0000	0,0529	0,0948	0,0892	0,1480

Tabela 5. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Olho azul		Olho verde		Olho mel		Olho castanho claro		Olho castanho escuro		Total	
		2n= 26	2n=74	2n= 44	2n=122	2n= 16	2n= 38	2n= 120	2n= 142	2n= 190	2n=290	2n= 396	2n= 666
rs531268453 ¹ (69788550)	C/C	1,0000	0,9730	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9970
	C/T	0,0000	0,0270	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0030
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	0,9865	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9985
	T	0,0000	0,0135	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	0,0270	-	-	-	-	-	-	-	-	-	0,0030
	<i>He</i>	-	0,0270	-	-	-	-	-	-	-	-	-	0,0030
	<i>pEHW</i>	-	1,0000	-	-	-	-	-	-	-	-	-	1,0000
rs114935806 ¹ (69788560)	C/C	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9862	1,0000	0,9940
	C/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0138	0,0000	0,0060
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9931	1,0000	0,9970
	T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0069	0,0000	0,0030
	<i>Ho</i>	-	-	-	-	-	-	-	-	-	0,0138	-	0,0060
	<i>He</i>	-	-	-	-	-	-	-	-	-	0,0138	-	0,0060
	<i>pEHW</i>	-	-	-	-	-	-	-	-	-	1,0000	-	1,0000
rs561386441 ¹ (69788574)	C/C	1,0000	1,0000	1,0000	1,0000	1,0000	0,9474	0,9833	0,9859	1,0000	0,9862	0,9949	0,9880
	C/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0526	0,0167	0,0141	0,0000	0,0138	0,0051	0,0120
	G/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	1,0000	1,0000	1,0000	1,0000	0,9737	0,9917	0,9930	1,0000	0,9931	0,9975	0,9940
	G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0263	0,0083	0,0070	0,0000	0,0069	0,0025	0,0060
	<i>Ho</i>	-	-	-	-	-	0,0526	0,0167	0,0141	-	0,0138	0,0051	0,0120
	<i>He</i>	-	-	-	-	-	0,0526	0,0167	0,0141	-	0,0138	0,0051	0,0120
	<i>pEHW</i>	-	-	-	-	-	1,0000	1,0000	1,0000	-	1,0000	1,0000	1,0000

Tabela 5. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Olho azul		Olho verde		Olho mel		Olho castanho claro		Olho castanho escuro		Total	
		2n= 26	2n=74	2n= 44	2n=122	2n= 16	2n= 38	2n= 120	2n= 142	2n= 190	2n=290	2n= 396	2n= 666
rs113085085 ² (69801287)	A/A	1,0000	0,9730	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9970
	A/C	0,0000	0,0270	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0030
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	0,9865	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9985
	C	0,0000	0,0135	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	0,0270	-	-	-	-	-	-	-	-	-	0,0030
	<i>He</i>	-	0,0270	-	-	-	-	-	-	-	-	-	0,0030
	<i>pEHW</i>	-	1,0000	-	-	-	-	-	-	-	-	-	1,0000
rs13322505 ² (69801975)	T/T	1,0000	1,0000	1,0000	0,9835	1,0000	1,0000	1,0000	1,0000	0,9579	0,9586	0,9798	0,9819
	T/C	0,0000	0,0000	0,0000	0,0164	0,0000	0,0000	0,0000	0,0000	0,0421	0,0414	0,0202	0,0181
	C/C	0,0000	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	T	1,0000	1,0000	1,0000	0,9917	1,0000	1,0000	1,0000	1,0000	0,9790	0,9793	0,9899	0,9895
	C	0,0000	0,0000	0,0000	0,0083	0,0000	0,0000	0,0000	0,0000	0,0210	0,0207	0,0101	0,0105
	<i>Ho</i>	-	-	-	-	-	-	-	-	0,0421	0,0414	0,0202	0,0181
	<i>He</i>	-	-	-	-	-	-	-	-	0,0414	0,0407	0,0201	0,0179
	<i>pEHW</i>	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs76992570 ² (69807750)	G/G	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9856	1,0000	0,9937
	G/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0144	0,0000	0,0063
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9929	1,0000	0,9969
	A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0071	0,0000	0,0031
	<i>Ho</i>	-	-	-	-	-	-	-	-	-	0,0143	-	0,0635
	<i>He</i>	-	-	-	-	-	-	-	-	-	0,0143	-	0,0634
	<i>pEHW</i>	-	-	-	-	-	-	-	-	-	1,0000	-	1,0000

Tabela 5. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Olho azul		Olho verde		Olho mel		Olho castanho claro		Olho castanho escuro		Total	
		2n= 26	2n=74	2n= 44	2n=122	2n= 16	2n= 38	2n= 120	2n= 142	2n= 190	2n=290	2n= 396	2n= 666
rs80163893 ² (69812860)	C/C	1,0000	1,0000	1,0000	0,9672	1,0000	1,0000	0,9833	0,9859	1,0000	1,0000	0,9949	0,9910
	C/T	0,0000	0,0000	0,0000	0,0328	0,0000	0,0000	0,0167	0,0141	0,0000	0,0000	0,0051	0,0090
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	1,0000	1,0000	0,9836	1,0000	1,0000	0,9917	0,9930	1,0000	1,0000	0,9975	0,9955
	T	0,0000	0,0000	0,0000	0,0164	0,0000	0,0000	0,0083	0,0070	0,0000	0,0000	0,0025	0,0045
	<i>Ho</i>	-	-	-	0,0328	-	-	0,0167	0,0141	-	-	0,0051	0,0090
	<i>He</i>	-	-	-	0,0325	-	-	0,0167	0,0141	-	-	0,0051	0,0090
	<i>pEHW</i>	-	-	-	1,0000	-	-	1,0000	1,0000	-	-	1,0000	1,0000
rs72950039 ² (69812903)	C/C	1,0000	1,0000	0,9545	0,9672	1,0000	1,0000	0,9500	0,9577	0,9789	0,9724	0,9697	0,9730
	C/T	0,0000	0,0000	0,0455	0,0328	0,0000	0,0000	0,0500	0,0423	0,0211	0,0276	0,0303	0,0270
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	1,0000	0,9773	0,9836	1,0000	1,0000	0,9750	0,9789	0,9895	0,9862	0,9849	0,9865
	T	0,0000	0,0000	0,0227	0,0164	0,0000	0,0000	0,0250	0,0211	0,0011	0,0138	0,0152	0,0135
	<i>Ho</i>	-	-	0,0455	0,0328	-	-	0,0500	0,0423	0,0211	0,0276	0,0303	0,0270
	<i>He</i>	-	-	0,0455	0,0325	-	-	0,0492	0,0417	0,0209	0,0273	0,0299	0,0267
	<i>pEHW</i>	-	-	1,0000	1,0000	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
rs2217456 ² (69887578)	A/A	0,0000	0,1176	0,2000	0,1930	0,3750	0,3158	0,1964	0,1818	0,3152	0,3643	0,2500	0,2658
	A/G	0,5833	0,5588	0,3000	0,3509	0,5000	0,4737	0,4107	0,4091	0,3913	0,4071	0,4043	0,4177
	G/G	0,4167	0,3235	0,5000	0,4561	0,1250	0,2105	0,3929	0,4091	0,2935	0,2286	0,3457	0,3165
	A	0,2917	0,3913	0,3659	0,3675	0,6250	0,5526	0,4052	0,3897	0,5135	0,5689	0,4555	0,4743
	G	0,7083	0,6087	0,6341	0,6325	0,3750	0,4474	0,5948	0,6103	0,4865	0,4311	0,5445	0,5257
	<i>Ho</i>	0,5833	0,5588	0,3000	0,3509	0,5000	0,4737	0,4107	0,4091	0,3913	0,4071	0,4043	0,4177
	<i>He</i>	0,4312	0,4860	0,4667	0,4695	0,5000	0,5078	0,4850	0,4778	0,5025	0,4926	0,4967	0,4995
	<i>pEHW</i>	0,4865	0,4798	0,1494	0,0850	1,0000	1,0000	0,2749	0,3018	0,0377	0,0567	0,0112	0,0048

Tabela 5. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Olho azul		Olho verde		Olho mel		Olho castanho claro		Olho castanho escuro		Total	
		2n= 26	2n=74	2n= 44	2n=122	2n= 16	2n= 38	2n= 120	2n= 142	2n= 190	2n=290	2n= 396	2n= 666
rs78124318 ² (69887807)	G/G	0,7692	0,8649	0,8182	0,8361	1,0000	0,9474	0,9167	0,9296	0,9579	0,9517	0,9192	0,9159
	G/A	0,2308	0,1351	0,1818	0,1639	0,0000	0,0526	0,0833	0,0704	0,0421	0,0483	0,0808	0,0841
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	0,8846	0,9324	0,9091	0,9180	1,0000	0,9737	0,9583	0,9648	0,9790	0,9759	0,9596	0,9580
	A	0,1154	0,0676	0,0909	0,0820	0,0000	0,0263	0,4167	0,0352	0,0211	0,0241	0,0404	0,0420
	<i>Ho</i>	0,2308	0,1351	0,1818	0,1639	-	0,0526	0,0833	0,0704	0,0421	0,0483	0,0808	0,0841
	<i>He</i>	0,2123	0,1277	0,1691	0,1517	-	0,0526	0,0805	0,0684	0,0414	0,0473	0,0777	0,0807
	<i>pEHW</i>	1,0000	1,0000	1,0000	1,0000	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
rs112598669 ² (69985718)	G/G	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9789	0,9655	0,9898	0,9849
	G/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0211	0,0345	0,0102	0,0151
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9895	0,9828	0,9949	0,9925
	A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0105	0,0172	0,0051	0,0075
	<i>Ho</i>	-	-	-	-	-	-	-	-	0,0211	0,0345	0,0102	0,0151
	<i>He</i>	-	-	-	-	-	-	-	-	0,0209	0,0340	0,0101	0,0150
	<i>pEHW</i>	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs576634776 ³ (70008502)	C/C	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9859	1,0000	1,0000	1,0000	0,9970
	C/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0141	0,0000	0,0000	0,0000	0,0030
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9930	1,0000	1,0000	1,0000	0,9985
	T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0070	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	-	-	-	-	-	-	0,0141	-	-	-	0,0030
	<i>He</i>	-	-	-	-	-	-	-	0,0141	-	-	-	0,0030
	<i>pEHW</i>	-	-	-	-	-	-	-	1,0000	-	-	-	1,0000

Tabela 5. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Olho azul		Olho verde		Olho mel		Olho castanho claro		Olho castanho escuro		Total	
		2n= 26	2n=74	2n= 44	2n=122	2n= 16	2n= 38	2n= 120	2n= 142	2n= 190	2n=290	2n= 396	2n= 666
70014320 ³	C/C	1,0000	0,9730	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9970
	C/T	0,0000	0,0270	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0030
	T/T	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	1,0000	0,9865	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9985
	T	0,0000	0,0135	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	0,0270	-	-	-	-	-	-	-	-	-	0,0030
	<i>He</i>	-	0,0270	-	-	-	-	-	-	-	-	-	0,0030
	<i>pEHW</i>	-	1,0000	-	-	-	-	-	-	-	-	-	1,0000
rs36118030 ³ (70014384)	G/G	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9895	0,9793	0,9949	0,9910
	G/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0105	0,0207	0,0051	0,0090
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9947	0,9897	0,9975	0,9955
	A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0053	0,0103	0,0025	0,0045
	<i>Ho</i>	-	-	-	-	-	-	-	-	0,0105	0,0207	0,0051	0,0090
	<i>He</i>	-	-	-	-	-	-	-	-	0,0105	0,0206	0,0051	0,0090
	<i>pEHW</i>	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs150314710 ⁴ (70015549)	T/T	0,9231	0,9459	1,0000	0,9836	1,0000	1,0000	0,9500	0,9577	0,9684	0,9586	0,9646	0,9640
	T/C	0,0769	0,0541	0,0000	0,0164	0,0000	0,0000	0,0500	0,0423	0,0316	0,0414	0,0354	0,0360
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	T	0,9615	0,9730	1,0000	0,9918	1,0000	1,0000	0,9750	0,9789	0,9842	0,9793	0,9823	0,9820
	C	0,3846	0,0270	0,0000	0,0082	0,0000	0,0000	0,0250	0,0211	0,0158	0,0207	0,0177	0,1802
	<i>Ho</i>	0,0769	0,0541	-	0,0164	-	-	0,0500	0,0423	0,0316	0,0414	0,0354	0,0360
	<i>He</i>	0,0769	0,0533	-	0,0164	-	-	0,0492	0,0417	0,0312	0,0407	0,0348	0,0354
	<i>pEHW</i>	1,0000	1,0000	-	1,0000	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabela 5. Frequências alélicas e genotípicas, heterozigose observada (H_o) e esperada (H_e), e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos olhos. Para cada característica, os valores referentes à amostra populacional aleatória estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de p significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(conclusão)

Ponto de variação		Olho azul		Olho verde		Olho mel		Olho castanho claro		Olho castanho escuro		Total	
		2n= 26	2n=74	2n= 44	2n=122	2n= 16	2n= 38	2n= 120	2n= 142	2n= 190	2n=290	2n= 396	2n= 666
rs2131025 ⁴ (70015647)	T/T	0,8462	0,7838	0,7273	0,8197	0,8750	0,9474	0,8667	0,8169	0,8526	0,8690	0,8434	0,8438
	T/C	0,1538	0,2162	0,2727	0,1803	0,1250	0,0526	0,1167	0,1690	0,1368	0,1241	0,1465	0,1502
	C/C	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0167	0,0141	0,0105	0,0069	0,0101	0,0060
	T	0,9231	0,8919	0,8636	0,9098	0,9375	0,9737	0,9250	0,9014	0,9211	0,9310	0,9167	0,9189
	C	0,0769	0,1081	0,1364	0,0902	0,0625	0,0263	0,0750	0,0986	0,0790	0,0690	0,0833	0,0811
	H_o	0,1539	0,2162	0,2727	0,1803	0,1250	0,0526	0,1167	0,1690	0,1368	0,1241	0,1465	0,1502
	H_e	0,1477	0,1955	0,2410	0,1654	0,1250	0,0526	0,1399	0,1790	0,1462	0,1289	0,1532	0,1492
	p_{EHW}	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,2763	0,5094	0,4502	0,5052	0,6282	1,0000
rs78240629 ⁴ (70016569)	A/A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9895	0,9793	0,9949	0,9910
	A/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0105	0,0207	0,0051	0,0090
	G/G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9947	0,9897	0,9975	0,9955
	G	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0053	0,0103	0,0025	0,0045
	H_o	-	-	-	-	-	-	-	-	0,0105	0,0207	0,0051	0,0090
	H_e	-	-	-	-	-	-	-	-	0,0105	0,0206	0,0051	0,0090
	p_{EHW}	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs77962238 ⁴ (70016857)	G/G	0,9231	0,9459	1,0000	0,9836	1,0000	1,0000	0,9500	0,9577	0,9684	0,9586	0,9646	0,9640
	G/A	0,0769	0,0541	0,0000	0,0164	0,0000	0,0000	0,0500	0,0423	0,0316	0,0414	0,0354	0,0360
	A/A	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	G	0,9615	0,9730	1,0000	0,9918	1,0000	1,0000	0,9750	0,9789	0,9842	0,9793	0,9823	0,9820
	A	0,0385	0,0270	0,0000	0,0082	0,0000	0,0000	0,0250	0,0211	0,0158	0,0207	0,0177	0,0180
	H_o	0,0769	0,0541	-	0,0164	-	-	0,0500	0,0423	0,0316	0,0414	0,0354	0,0360
	H_e	0,0769	0,0533	-	0,0164	-	-	0,0492	0,0417	0,0312	0,0407	0,0348	0,0354
	p_{EHW}	1,0000	1,0000	-	1,0000	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabela 6. Frequências alélicas e genotípicas, heterozigose observada (H_o) e esperada (H_e), e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos. Para cada característica, os valores referentes à amostra populacional estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de p significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continua)

Ponto de variação		Cabelo ruivo		Cabelo loiro claro		Cabelo loiro escuro		Cabelo castanho claro		Cabelo castanho escuro		Cabelo preto		Total	
		2n=0	2n=22	2n=4	2n=26	2n=42	2n=68	2n=78	2n=120	2n=172	2n=244	2n=96	2n=170	2n=396	2n=666
rs74845300 ¹ (69786714)	A/A	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9792	0,9294	0,9949	0,9820
	A/G	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0208	0,0706	0,0051	0,0180
	G/G	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9896	0,9647	0,9975	0,9910
	G	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0104	0,0353	0,0025	0,0090
	H_o	-	-	-	-	-	-	-	-	-	-	0,0208	0,0706	0,0051	0,0180
	H_e	-	-	-	-	-	-	-	-	-	-	0,0208	0,0685	0,0051	0,0179
	p_{EHW}	-	-	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs80232752 ¹ (69787573)	A/A	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9792	0,9294	0,9949	0,9820
	A/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0208	0,0706	0,0051	0,0180
	T/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9896	0,9647	0,9975	0,9910
	T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0104	0,0353	0,0025	0,0090
	H_o	-	-	-	-	-	-	-	-	-	-	0,0208	0,0706	0,0051	0,0180
	H_e	-	-	-	-	-	-	-	-	-	-	0,0208	0,0685	0,0051	0,0179
	p_{EHW}	-	-	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000
rs12489786 ¹ (69788068)	T/T	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9833	0,9186	0,9262	0,9792	0,9529	0,9596	0,9580
	T/C	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0167	0,0698	0,0656	0,0208	0,0471	0,0354	0,0390
	C/C	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0116	0,0082	0,0000	0,0000	0,0051	0,0030
	T	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9917	0,9535	0,9590	0,9896	0,9765	0,9773	0,9775
	C	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0083	0,0465	0,0410	0,0104	0,0235	0,0227	0,0225
	H_o	-	-	-	-	-	-	-	0,0167	0,0698	0,0656	0,0208	0,0471	0,0354	0,0390
	H_e	-	-	-	-	-	-	-	0,0167	0,0892	0,0789	0,0208	0,0462	0,0445	0,0441
	p_{EHW}	-	-	-	-	-	-	-	1,0000	0,1572	0,1748	1,0000	1,0000	0,0892	0,1480

Tabela 6. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos. Para cada característica, os valores referentes à amostra populacional estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Cabelo ruivo		Cabelo loiro claro		Cabelo loiro escuro		Cabelo castanho claro		Cabelo castanho escuro		Cabelo preto		Total	
		2n=0	2n=22	2n=4	2n=26	2n=42	2n=68	2n=78	2n=120	2n=172	2n=244	2n=96	2n=170	2n=396	2n=666
rs531268453 ¹ (69788550)	C/C	-	1,0000	1,0000	0,9231	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9970
	C/T	-	0,0000	0,0000	0,0769	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0030
	T/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	-	1,0000	1,0000	0,9615	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9985
	T	-	0,0000	0,0000	0,0385	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0015
	<i>Ho</i>	-	-	-	0,0769	-	-	-	-	-	-	-	-	-	0,0030
	<i>He</i>	-	-	-	0,0769	-	-	-	-	-	-	-	-	-	0,0030
	<i>pEHW</i>	-	-	-	1,0000	-	-	-	-	-	-	-	-	-	1,0000
rs114935806 ¹ (69788560)	C/C	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9765	1,0000	0,9940
	C/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0235	0,0000	0,0060
	T/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9882	1,0000	0,9970
	T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0118	0,0000	0,0030
	<i>Ho</i>	-	-	-	-	-	-	-	-	-	-	-	0,0235	-	0,0060
	<i>He</i>	-	-	-	-	-	-	-	-	-	-	-	0,0234	-	0,0060
	<i>pEHW</i>	-	-	-	-	-	-	-	-	-	-	-	1,0000	-	1,0000
rs561386441 ¹ (69788574)	C/C	-	1,0000	1,0000	1,0000	1,0000	0,9420	1,0000	1,0000	0,9884	0,9836	1,0000	0,9765	0,9949	0,9880
	C/G	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0116	0,0164	0,0000	0,0235	0,0051	0,0120
	G/G	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	-	1,0000	1,0000	1,0000	1,0000	0,9706	1,0000	1,0000	0,9942	0,9918	1,0000	0,9882	0,9975	0,9940
	G	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0058	0,0082	0,0000	0,0118	0,0025	0,0060
	<i>Ho</i>	-	-	-	-	-	-	-	-	0,0116	0,0164	-	0,0235	0,0051	0,0120
	<i>He</i>	-	-	-	-	-	-	-	-	0,0116	0,0163	-	0,0234	0,0051	0,0120
	<i>pEHW</i>	-	-	-	-	-	-	-	-	1,0000	1,0000	-	1,0000	1,0000	1,0000

Tabela 6. Frequências alélicas e genotípicas, heterozigose observada (H_o) e esperada (H_e), e aderência ao Equilíbrio de Hardy-Weinberg (p_{EHW}) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos. Para cada característica, os valores referentes à amostra populacional estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de p significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Cabelo ruivo		Cabelo loiro claro		Cabelo loiro escuro		Cabelo castanho claro		Cabelo castanho escuro		Cabelo preto		Total	
		2n=0	2n=22	2n=4	2n=26	2n=42	2n=68	2n=78	2n=120	2n=172	2n=244	2n=96	2n=170	2n=396	2n=666
rs113085085 ² (69801287)	A/A	-	1,0000	1,0000	0,9231	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9970
	A/C	-	0,0000	0,0000	0,0769	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0030
	C/C	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	A	-	1,0000	1,0000	0,9615	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9985
	C	-	0,0000	0,0000	0,0385	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0015
	H_o	-	-	-	0,0769	-	-	-	-	-	-	-	-	-	0,0030
	H_e	-	-	-	0,0769	-	-	-	-	-	-	-	-	-	0,0030
	p_{EHW}	-	-	-	1,0000	-	-	-	-	-	-	-	-	-	1,0000
rs13322505 ² (69801975)	T/T	-	1,0000	1,0000	1,0000	1,0000	0,9704	1,0000	1,0000	0,9884	0,9918	0,9375	0,9412	0,9798	0,9819
	T/C	-	0,0000	0,0000	0,0000	0,0000	0,0294	0,0000	0,0000	0,0116	0,0082	0,0625	0,0588	0,0202	0,0181
	C/C	-	0,0000	0,0000	0,0000	0,0000	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	T	-	1,0000	1,0000	1,0000	1,0000	0,9851	1,0000	1,0000	0,9942	0,9959	0,9688	0,9706	0,9899	0,9895
	C	-	0,0000	0,0000	0,0000	0,0000	0,0149	0,0000	0,0000	0,0058	0,0041	0,0313	0,2941	0,0101	0,0105
	H_o	-	-	-	-	-	-	-	-	0,0116	0,0082	0,0625	0,0588	0,0202	0,0181
	H_e	-	-	-	-	-	-	-	-	0,0116	0,0082	0,6118	0,5743	0,0201	0,0179
	p_{EHW}	-	-	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
rs76992570 ² (69807750)	G/G	-	1,0000	1,0000	1,0000	0,9070	1,0000	0,9246	1,0000	0,9315	0,9915	0,9588	0,9877	1,0000	0,9937
	G/A	-	0,0000	0,0000	0,0000	0,0907	0,0000	0,0740	0,0000	0,0673	0,0085	0,0408	0,0123	0,0000	0,0063
	A/A	-	0,0000	0,0000	0,0000	0,0023	0,0000	0,0015	0,0000	0,0012	0,0000	0,0004	0,0000	0,0000	0,0000
	G	-	1,0000	1,0000	1,0000	0,9524	1,0000	0,9615	1,0000	0,9651	0,9958	0,9792	0,9940	1,0000	0,9969
	A	-	0,0000	0,0000	0,0000	0,0476	0,0000	0,0385	0,0000	0,0349	0,0042	0,0208	0,0060	0,0000	0,0031
	H_o	-	-	-	-	-	-	-	-	-	0,0085	-	0,0124	-	0,0635
	H_e	-	-	-	-	-	-	-	-	-	0,0085	-	0,0124	-	0,0634
	p_{EHW}	-	-	-	-	-	-	-	-	-	1,0000	-	1,0000	-	1,0000

Tabela 6. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos. Para cada característica, os valores referentes à amostra populacional estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

Ponto de variação		Cabelo ruivo		Cabelo loiro claro		Cabelo loiro escuro		Cabelo castanho claro		Cabelo castanho escuro		Cabelo preto		Total	
		2n=0	2n=22	2n=4	2n=26	2n=42	2n=68	2n=78	2n=120	2n=172	2n=244	2n=96	2n=170	2n=396	2n=666
rs80163893 ² (69812860)	C/C	-	1,0000	1,0000	1,0000	1,0000	1,0000	0,9744	0,9500	1,0000	1,0000	1,0000	1,0000	0,9949	0,9910
	C/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0256	0,0500	0,0000	0,0000	0,0000	0,0000	0,0051	0,0090
	T/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	-	1,0000	1,0000	1,0000	1,0000	1,0000	0,9872	0,9750	1,0000	1,0000	1,0000	1,0000	0,9975	0,9955
	T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0128	0,0250	0,0000	0,0000	0,0000	0,0000	0,0025	0,0045
	<i>Ho</i>	-	-	-	-	-	-	0,0256	0,0500	-	-	-	-	0,0051	0,0090
	<i>He</i>	-	-	-	-	-	-	0,0256	0,0492	-	-	-	-	0,0051	0,0090
	<i>pEHW</i>	-	-	-	-	-	-	1,0000	1,0000	-	-	-	-	1,0000	1,0000
rs72950039 ² (69812903)	C/C	-	1,0000	1,0000	1,0000	1,0000	1,0000	0,9487	0,9667	0,9535	0,9508	1,0000	0,9882	0,9697	0,9730
	C/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0513	0,0333	0,0465	0,0492	0,0000	0,0118	0,0303	0,0270
	T/T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	C	-	1,0000	1,0000	1,0000	1,0000	1,0000	0,9744	0,9833	0,9767	0,9754	1,0000	0,9941	0,9849	0,9865
	T	-	0,0000	0,0000	0,0000	0,0000	0,0000	0,0256	0,0167	0,0233	0,0246	0,0000	0,0059	0,0152	0,0135
	<i>Ho</i>	-	-	-	-	-	-	0,0513	0,0333	0,0465	0,0492	-	0,0118	0,0303	0,0270
	<i>He</i>	-	-	-	-	-	-	0,0506	0,0331	0,0457	0,0482	-	0,0118	0,0299	0,0267
	<i>pEHW</i>	-	-	-	-	-	-	1,0000	1,0000	1,0000	1,0000	-	1,0000	1,0000	1,0000
rs2217456 ² (69887578)	A/A	-	0,1000	0,0000	0,1667	0,1053	0,1000	0,2222	0,2143	0,2410	0,2605	0,3696	0,4074	0,2500	0,2658
	A/G	-	0,5000	1,0000	0,5000	0,3158	0,3000	0,4167	0,5179	0,4458	0,4370	0,3043	0,3457	0,4043	0,4177
	G/G	-	0,4000	0,0000	0,3333	0,5789	0,6000	0,3611	0,2679	0,3133	0,3025	0,3261	0,2469	0,3457	0,3165
	A	-	0,3333	0,5000	0,4167	0,2632	0,2419	0,4400	0,4741	0,4643	0,4792	0,5269	0,5854	0,4555	0,4743
	G	-	0,6667	0,5000	0,5833	0,7368	0,7581	0,5600	0,5259	0,5357	0,5208	0,4731	0,4146	0,5445	0,5257
	<i>Ho</i>	-	0,5000	1,0000	0,5000	0,3158	0,3000	0,4167	0,5179	0,4458	0,4370	0,3044	0,3457	0,4043	0,4177
	<i>He</i>	-	0,4790	0,6667	0,5073	0,3983	0,3814	0,4973	0,5031	0,5004	0,5012	0,5045	0,4902	0,4967	0,4995
	<i>pEHW</i>	-	1,0000	1,0000	1,0000	0,5503	0,3285	0,4950	1,0000	0,3803	0,1985	0,0082	0,0116	0,0112	0,0048

Tabela 6. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos. Para cada característica, os valores referentes à amostra populacional estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

[illegible]

Tabela 6. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos. Para cada característica, os valores referentes à amostra populacional estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continuação)

[illegible]

Tabela 6. Frequências alélicas e genotípicas, heterozigose observada (*Ho*) e esperada (*He*), e aderência ao Equilíbrio de Hardy-Weinberg (*pEHW*) dos 21 pontos de variação associados a algum fenótipo de pigmentação para indivíduos agrupados segundo a cor dos cabelos. Para cada característica, os valores referentes à amostra populacional estão à esquerda e os referentes à amostragem total estão à direita. *Singletons* e valores de *p* significativos estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(conclusão)

[illegible]

6.4.1. Promotora

Foram encontrados 17 haplótipos para a região promotora considerando 18 pontos de variação, como mostrado na Tabela 7. Os haplótipos foram resolvidos com probabilidade maior que 80% para 329 dos 333 indivíduos não aparentados sequenciados. Como dito anteriormente, um desses pontos se mostrou monomórfico dentre esses 329 indivíduos, restando 17 pontos de variação. Os haplótipos foram nomeados levando em consideração a frequência e a relação molecular entre eles, como mostra a Figura 6. A diversidade haplotípica para essa região foi de $0,7908 \pm 0,0096$.

Figura 6. Relação entre os 17 haplótipos encontrados para a região promotora do gene *MITF*. O tamanho dos círculos representa a frequência do haplótipo na população ($n = 329$) e o número de pontos nos ramos que conectam dois haplótipos representa o número de diferenças entre eles. Haplótipos associados a fenótipos escuros estão coloridos em marrom escuro.

Tabela 7. Haplótipos encontrados para região promotora considerando 17 pontos de variação em 329 amostras. Variantes e haplótipos associados à algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipo	rs200471673	rs62253052	rs9824825	rs9837720	rs181456408	rs57953945	rs139094245	rs74845300	rs182103206	rs150351819	rs80232752	rs1427421	rs12489786	rs7651891	rs114935806	rs69788567	rs561386441	Frequência
Promo-1	AT	T	T	G	G	CT	CATGTTTCTCAATGGGG	A	G	C	A	G	T	G	C	C	C	0,3571
Promo-1a	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,2128
Promo-1a1	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	0,0030
Promo-1b	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	0,1337
Promo-1b1	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	0,1018
Promo-1b1a	.	.	.	.	A	.	.	.	.	.	.	A	.	A	.	.	.	0,0061
Promo-1b2	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	0,0821
Promo-1b2a	.	.	C	A	.	.	.	.	.	.	.	.	.	A	.	.	.	0,0076
Promo-1b3	TT	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	0,0046
Promo-1c	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	0,0350
Promo-1d	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	0,0167
Promo-1d1	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	0,0061
Promo-1e	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0152
Promo-1f	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	0,0030
Promo-1g	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	0,0030
Promo-1h1	.	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	0,0091
Promo-1i1	.	.	.	.	.	C	.	.	.	.	.	.	.	.	T	.	.	0,0030

6.4.2. 5'UTR e CD

Na extensão da região 5'UTR e CD foram encontrados 50 haplótipos considerando 38 pontos de variação, como mostra a Tabela 8. Os haplótipos dessa região foram resolvidos com probabilidade maior que 80% para 307 indivíduos e foram nomeados conforme a frequência e a relação molecular entre eles, representada na Figura 7. Como dito anteriormente, dois pontos de variação se mostraram monomórficos dentre esses 307 indivíduos, restando 36 pontos de variação. A diversidade haplotípica para essa região foi de $0,8072 \pm 0,0113$.

Figura 7. Relação entre os 50 haplótipos encontrados para a região 5'UTR e CD do gene *MITF*. O tamanho dos círculos representa a frequência do haplótipo na população ($n = 307$) e o número de pontos nos ramos que conectam dois haplótipos representa o número de diferenças entre eles. Haplótipos associados a fenótipos intermediários e escuros estão coloridos em marrom claro e escuro, respectivamente.

Tabela 8. Haplótipos encontrados para região 5'UTR e codificante considerando 36 pontos de variação em 307 amostras. Variantes e haplótipos associados à algum fenótipo de pigmentação estão destacados em vermelho.

(continua)

Haplótipo	rs530113723	rs6761994	rs13322505	rs72950030	rs76992570	rs4855323	rs62250968	rs72950037	rs80163893	rs72950039	rs2217456	rs1367370	rs1367371	rs545897666	rs78124318	rs9849776	rs112598669	rs77588960	rs189927456	rs73117325	rs113394882	rs78545229	rs145438735	rs113310138	rs188710184	chr3:69987914	rs374067688	rs28583802	rs79668299	chr3:69998393	chr3:69998442	rs186002522	chr3:70000957	rs55754687	rs114827487	rs36118030	Freq		
5'CD-1	A	G	T	C	G	G	C	G	C	C	A	A	C	T	G	G	G	G	T	A	C	C	G	C	T	A	T	A	C	G	T	C	G	T	C	G	0,351		
5'CD-1a	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,042	
5'CD-1a1	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,004	
5'CD-1b	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	0,013	
5'CD-1c	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,011	
5'CD-1d	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,011	
5'CD-1e	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,004	
5'CD-1e1a	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	0,004	
5'CD-1e2a	.	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	0,003	
5'CD-1f	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,004
5'CD-1f1	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,003
5'CD-1f1a1	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	0,001	
5'CD-1g	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,003
5'CD-1g1	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,009
5'CD-1h	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,003
5'CD-1i	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	0,003
5'CD-1j	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,003
5'CD-1k1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,004
5'CD-2	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,236
5'CD-2a	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,019
5'CD-2b	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,016
5'CD-2c	.	.	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,008
5'CD-2c1	.	.	.	.	.	.	T	.	.	.	G	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,006
5'CD-2c2	.	.	C	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,003

Tabela 8. Haplótipos encontrados para região 5'UTR e codificante considerando 36 pontos de variação em 307 amostras. Variantes e haplótipos associados à algum fenótipo de pigmentação estão destacados em vermelho.

(continuação)

Haplótipo	rs530113723	rs6761994	rs13322505	rs72950030	rs76992570	rs4855323	rs62250968	rs72950037	rs80163893	rs72950039	rs2217456	rs1367370	rs1367371	rs545897666	rs78124318	rs9849776	rs112598669	rs77588960	rs189927456	rs73117325	rs113394882	rs78545229	rs145438735	rs113310138	rs188710184	chr3:69987914	rs374067688	rs28583802	rs79668299	chr3:69998393	chr3:69998442	rs186002522	chr3:70000957	rs55754687	rs114827487	rs36118030	Freq	
5'CD-1	A	G	T	C	G	G	C	G	C	C	A	A	C	T	G	G	G	G	T	A	C	C	G	C	T	A	T	A	C	G	T	C	G	T	C	G	0,3518	
5'CD-2c2b	.	.	C	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	0,0049
5'CD-2d	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0049
5'CD-2e	.	A	.	.	.	.	T	.	.	.	G	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0049
5'CD-2f	.	A	.	.	.	.	T	.	.	.	G	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0033
5'CD-2g	.	A	.	.	.	.	T	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0033
5'CD-2h	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	0,0033
5'CD-2i	.	A	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0016
5'CD-2i	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0016
5'CD-2k1	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0065
5'CD-2k1a1	.	.	.	.	.	.	T	.	.	.	G	G	A	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0049
5'CD-2l1	.	A	.	.	.	.	T	.	.	.	G	.	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	0,0049
5'CD-2l1a	.	.	.	.	.	.	T	.	.	.	G	.	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	0,0016
5'CD-3	.	.	.	.	.	T	.	.	.	.	G	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,1059
5'CD-3a	.	.	.	.	.	T	.	.	.	.	G	G	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,013
5'CD-3a1	.	.	.	.	.	T	.	.	.	.	G	G	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	0,0195
5'CD-3a2a	.	.	.	.	.	T	.	.	.	.	G	G	A	.	A	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0081
5'CD-3b	.	.	.	.	.	T	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0098
5'CD-3c	.	.	.	.	.	.	.	.	.	.	G	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0065
5'CD-3c1a1	.	.	.	.	.	.	.	.	.	.	G	G	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	0,0033
5'CD-3c1a1a	.	.	.	.	.	.	.	.	.	T	G	G	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	0,0016
5'CD-3c1a1a1	.	.	.	.	.	.	.	.	.	T	G	G	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	0,0016
5'CD-3c1a1a1a	.	.	.	.	.	.	.	.	.	T	.	G	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	0,0016
5'CD-3d	.	.	.	.	.	T	T	.	.	.	G	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0033

(conclusão)

[illegible]

6.4.3. 3'UTR

Na região 3'UTR foram encontrados 18 haplótipos considerando 17 pontos de variação, como mostra a Tabela 9. Os haplótipos dessa região foram resolvidos com probabilidade maior que 80% para 329 indivíduos e foram nomeados conforme a frequência e a relação molecular entre eles, representada na Figura 8. Como dito anteriormente, dois pontos de variação se mostraram monomórficos dentre esses 329 indivíduos, restando 15 pontos de variação. A diversidade haplotípica para essa região foi de $0,6014 \pm 0,0178$.

Figura 8. Relação entre os 18 haplótipos encontrados para a região 3'UTR do gene *MITF*. O tamanho dos círculos representa a frequência do haplótipo na população ($n = 329$) e o número de pontos nos ramos que conectam dois haplótipos representa o número de diferenças entre eles. O haplótipo associado a fenótipo escuro está colorido em marrom escuro.

Tabela 9. Haplótipos encontrados para região 3'UTR considerando 15 pontos de variação em 329 amostras. Variantes e haplótipos associados à algum fenótipo de pigmentação estão destacados em vermelho.

Haplótipo	rs9863910	rs552044239	chr3:70014955	rs146187921	rs114659118	rs150314710	rs2131025	rs144336708	rs576	rs78240629	rs77962238	rs704246	rs139487027	chr3:70017154	rs151043552	Frequência
3UTR-1	T	T	T	G	T	T	T	G	C	A	G	C	C	G	C	0,5851
3UTR-1a	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0,0471
3UTR-1a1	C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	0,0669
3UTR-1a1a	C	.	C	.	.	.	C	.	.	.	.	.	.	.	.	0,0076
3UTR-1b	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	0,0106
3UTR-1c	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	0,0076
3UTR-1d	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	0,0061
3UTR-1d1	.	.	.	.	.	.	.	.	T	G	.	.	.	.	.	0,0030
3UTR-1d2a	.	.	.	.	.	C	.	.	T	.	A	.	.	.	.	0,0106
3UTR-1d2a1	.	.	.	.	.	C	.	A	T	.	A	.	.	.	.	0,0076
3UTR-1e	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	0,0046
3UTR-1f	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	0,0030
3UTR-1g	.	.	.	GAA	.	.	.	.	.	.	.	.	.	.	.	0,0015
3UTR-2	.	.	.	GAA	.	.	.	.	T	.	.	T	.	.	.	0,2204
3UTR-2a	.	.	.	.	.	.	.	.	T	.	.	T	.	.	.	0,0091
3UTR-2b	.	C	.	GAA	.	.	.	.	T	.	.	T	.	.	.	0,0046
3UTR-2c	.	.	.	GAA	.	.	.	.	T	.	.	T	.	A	.	0,0030
3UTR-2d	.	.	.	GAA	.	.	.	.	.	.	.	T	.	.	.	0,0015

6.4.4. MITF

Foram identificados 132 haplótipos para o gene *MITF* considerando 68 (17 na promotora, 36 na região 5'UTR e CD e 15 na região 3'UTR) pontos de variação em toda a extensão do gene, incluindo sua região promotora. Os haplótipos foram representados conforme a composição das regiões promotora, 5'UTR e CD e 3'UTR, e foram nomeados de acordo com a frequência. Foram identificados os haplótipos completos com probabilidade maior que 80% de 244 indivíduos. Dentre esses 244 indivíduos, mais três variantes se mostraram monomórficas. A diversidade haplotípica para essa região foi de $0,9781 \pm 0,0020$.

Na Tabela 10 (Apêndice B) estão representados os 132 haplótipos encontrados para o gene *MITF*, conforme sua composição haplotípica das regiões promotora, 5'UTR e CD e 3'UTR. Devido ao grande número de haplótipos encontrados para a extensão completa do gene

MITF e a complexidade da network formada por eles, na Figura 9 estão representados somente os haplótipos que apresentaram frequência acima de 1% nessa população.

Tabela 10. Haplótipos encontrados para a extensão total do gene *MITF* com frequência mínima de 1% considerando 68 pontos de variação em 244 amostras. Haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho. A tabela completa está representada no Apêndice B.

Haplótipo	Promotora	5'UTR + CD	3'UTR	Frequência
MITF-1	Promo-1	5'CD-2	3UTR-1	0,0738
MITF-2	Promo-1a	5'CD-2	3UTR-1	0,0512
MITF-3	Promo-1	5'CD-1a	3UTR-1	0,0471
MITF-4	Promo-1b	5'CD-1	3UTR-1	0,0451
MITF-5	Promo-1	5'CD-1	3UTR-1	0,0410
MITF-6	Promo-1b1	5'CD-1	3UTR-1a	0,0389
MITF-7	Promo-1a	5'CD-1	3UTR-1	0,0328
MITF-8	Promo-1b2	5'CD-1	3UTR-1	0,0328
MITF-9	Promo-1	5'CD-2	3UTR-1a1	0,0307
MITF-10	Promo-1	5'CD-3	3UTR-1	0,0307
MITF-11	Promo-1	5'CD-1	3UTR-2	0,0225
MITF-12	Promo-1	5'CD-3	3UTR-2	0,0184
MITF-13	Promo-1b	5'CD-2	3UTR-1	0,0143
MITF-14	Promo-1c	5'CD-3a1	3UTR-1	0,0143
MITF-15	Promo-1a	5'CD-2a	3UTR-1a1	0,0123
MITF-16	Promo-1b1	5'CD-2b	3UTR-1a1	0,0123
MITF-17	Promo-1a	5'CD-3	3UTR-1	0,0102
MITF-18	Promo-1b	5'CD-3	3UTR-2	0,0102
MITF-19	Promo-1b1	5'CD-1c	3UTR-1	0,0102
MITF-20	Promo-1b3	5'CD-1	3UTR-2	0,0102
MITF-21	Promo-1d	5'CD-1	3UTR-1	0,0102
MITF-22	Promo-1e	5'CD-1	3UTR-2	0,0102
MITF-23	Promo-1g	5'CD-2	3UTR-1	0,0102
MITF-24	Promo-1h1	5'CD-1	3UTR-2	0,0102

Figura 9. Relação entre os 24 haplótipos encontrados para o gene *MITF* com frequência acima de 1%. O tamanho dos círculos representa a frequência do haplótipo na população ($n = 244$) e os nós entre os haplótipos representam o número de diferenças entre eles. Haplótipos associados a fenótipos escuros, intermediários e claros estão coloridos em marrom escuro, marrom claro e bege, respectivamente.

6.5. Desequilíbrio de Ligação

Primeiramente o desequilíbrio de ligação foi avaliado nas regiões do gene *MITF* separadamente.

Na região promotora foram avaliados 16 pontos de variação bialélicos. Nessa região se formou um único bloco em desequilíbrio envolvendo 13 variantes, como mostra a Figura 10.

Figura 10. Representação do desequilíbrio de ligação na região promotora do gene *MITF*. O bloco em desequilíbrio está delimitado em preto. Vermelho indica $LOD (Log\ of\ the\ Odds) \geq 2$ e $D' = 1$; azul indica $LOD < 2$ e $D' = 1$; tons de rosa indicam $LOD \geq 2$ e $D' < 1$; branco indica $LOD < 2$ e $D' < 1$. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002) No canto inferior direito estão as frequências dos haplótipos compostos pelas variantes destacadas em negrito na parte superior da imagem (*tag* SNPs).

Na região da 5'UTR e CD foram avaliados 36 pontos de variação bialélicos. Nessa região foram encontrados dois blocos em desequilíbrio, compostos por seis e três pontos de variação, como mostra a Figura 11.

Figura 11. Representação do desequilíbrio de ligação na região 5'UTR e CD do gene *MITF*. Os blocos em desequilíbrio estão delimitados em preto. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). No canto inferior direito estão as frequências dos haplótipos compostos pelas variantes destacadas em negrito na parte superior da imagem (*tag* SNPs). Os traços entre os haplótipos de cada bloco, acompanhado do valor de D' entre os blocos, indicam a intensidade da ligação entre eles.

Na região 3'UTR foram avaliados 14 pontos de variação bialélicos e foi encontrado somente um bloco em desequilíbrio composto por seis pontos de variação, como mostrado na Figura 12.

Figura 12. Representação do desequilíbrio de ligação na região 3'UTR do gene *MTF*. O bloco em desequilíbrio está delimitado em preto. Vermelho indica $\text{LOD (Log of the Odds)} \geq 2$ e $D' = 1$; azul indica $\text{LOD} < 2$ e $D' = 1$; tons de rosa indicam $\text{LOD} \geq 2$ e $D' < 1$; branco indica $\text{LOD} < 2$ e $D' < 1$. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). No canto inferior direito estão as frequências dos haplótipos compostos pelas variantes destacadas em negrito na parte superior da imagem (*tag* SNPs).

A Figura 13 mostra o desequilíbrio de ligação na extensão total do gene *MTF*, incluindo sua região promotora, avaliando 63 pontos de variação bialélicos, uma vez que, para os 244 indivíduos que tiveram seus haplótipos inferidos com probabilidade de pelo menos 80%, 3 variantes (chr3:69788643, rs72950030 e rs145438735) presentes na região codificante se mostraram monomórficas. Essa figura mostra como os blocos presentes em cada uma das regiões do gene interagem e qual o nível de desequilíbrio (D') entre eles.

Figura 13. Representação do desequilíbrio de ligação na extensão total do gene *MTF*. Os blocos em desequilíbrio estão cercados em preto. Os blocos de haplótipos foram definidos de acordo com o método de Gabriel et al. (2002). No canto inferior direito estão as frequências dos haplótipos compostos pelas variantes destacadas em negrito na parte superior da imagem (*tag* SNPs).

6.6. Associações fenotípicas

Dos 133 pontos de variação analisados, 21 apresentam associação forte ($OR > 10$) e significativa ($p < 0,05$) com pelo menos um dos fenótipos de pigmentação, como mostrado na Tabela 11. Dessas 21 variantes, seis estão presentes na região promotora, onze na região codificante, sendo 8 em íntrons e 3 em éxons, e quatro variantes na região 3'UTR. Se a conservadora correção de Bonferroni para múltiplos testes for considerada, somente as associações das variantes rs74845300*G e rs80232752*T com cabelos pretos, seriam significativas.

Tabela 11. Associações fenotípicas encontradas para 21 pontos de variação dentre os 133 analisados em toda a extensão do gene *MITF*. *Singletons* estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente. Associações significativas mesmo após aplicação da correção de Bonferroni estão em negrito, considerando $\alpha_{\text{corrigido}} = 0,0004$ (isto é, $0,05/133$ SNPs).

(continua)

Ponto de variação	Alelo / Genótipo	Características	Valor de p	Odds ratio (95% IC)
rs74845300 ¹ (69786714)	A	Cabelo castanho escuro x preto	0,0041	20,2670 (1,1263-364,680)
	A/A	Cabelo castanho escuro x preto	0,0043	19,5585 (1,0945-349,506)
	G	Cabelo preto x não preto	0,0002	41,0993 (2,2902-737,5484)
	A/G	Cabelo preto x não preto	0,0002	39,7027 (2,2247-708,5203)
	G	Cabelo preto x castanho escuro	0,0041	20,2670 (1,1263-364,6807)
	A/G	Cabelo preto x castanho escuro	0,0043	19,5585 (1,0945-349,5064)
	G	Olho castanho escuro x não castanho escuro	0,0066	17,3484 (0,9694-310,4651)
	A/G	Olho castanho escuro x não castanho escuro	0,0068	17,0000 (0,9538-302,9839)
	A	Pele clara x escura	0,0015	30,8640 (1,6771-567,9725)
	A/A	Pele clara x escura	0,0016	29,5209 (1,6186-538,4082)
	G	Pele escura x não escura	0,0005	27,7450 (3,1753-242,4257)
	A/G	Pele escura x não escura	0,0005	26,4953 (3,0649-229,0446)
	G	Pele escura x média	0,0077	13,6274 (1,5545-119,4578)
	A/G	Pele escura x média	0,0082	13,0373 (1,5056-112,8873)
	G	Pele escura x clara	0,0015	30,8640 (1,6771-567,9725)
	A/G	Pele escura x clara	0,0016	29,5209 (1,6186-538,4082)
	A	Pele média x escura	0,0077	13,6274 (1,5545-119,4578)
	A/A	Pele média x escura	0,0082	13,0373 (1,5056-112,8873)
rs80232752 ¹ (69787573)	A	Cabelo castanho escuro x preto	0,0041	20,2670 (1,1263-364,680)
	A/A	Cabelo castanho escuro x preto	0,0043	19,5585 (1,0945-349,506)
	T	Cabelo preto x não preto	0,0002	41,0993 (2,2902-737,5484)
	A/T	Cabelo preto x não preto	0,0002	39,7027 (2,2247-708,5203)
	T	Cabelo preto x castanho escuro	0,0041	20,2670 (1,1263-364,6807)
	A/T	Cabelo preto x castanho escuro	0,0043	19,5585 (1,0945-349,5064)
	T	Olho castanho escuro x não castanho escuro	0,0066	17,3484 (0,9694-310,4651)
	A/T	Olho castanho escuro x não castanho escuro	0,0068	17,0000 (0,9538-302,9839)

Tabela 11. Associações fenotípicas encontradas para 21 pontos de variação dentre os 133 analisados em toda a extensão do gene *MITF*. *Singletons* estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente. Associações significativas mesmo após aplicação da correção de Bonferroni estão em negrito, considerando $\alpha_{\text{corrigido}} = 0,0004$ (isto é, $0,05/133$ SNPs).

(continuação)

Ponto de variação	Alelo / Genótipo	Características	Valor de <i>p</i>	Odds ratio (95% IC)
rs80232752¹ (69787573)	A	Pele clara x escura	0,0015	30,8640 (1,6771-567,9725)
	A/A	Pele clara x escura	0,0016	29,5209 (1,6186-538,4082)
	T	Pele escura x não escura	0,0005	27,7450 (3,1753-242,4257)
	A/T	Pele escura x não escura	0,0005	26,4953 (3,0649-229,0446)
	T	Pele escura x média	0,0077	13,6274 (1,5545-119,4578)
	A/T	Pele escura x média	0,0082	13,0373 (1,5056-112,8873)
	T	Pele escura x clara	0,0015	30,8640 (1,6771-567,9725)
	A/T	Pele escura x clara	0,0016	29,5209 (1,6186-538,4082)
	A	Pele média x escura	0,0077	13,6274 (1,5545-119,4578)
rs12489786¹ (69788068)	A/A	Pele média x escura	0,0082	13,0373 (1,5056-112,8873)
	T	Pele clara x média	0,0095	9,8244 (1,2279-78,6029)
	T/T	Pele clara x média	0,0053	10,6296 (1,3516-83,5945)
	T	Pele clara x escura	0,0223	11,0000 (1,2016-100,6921)
	T/T	Pele clara x escura	0,0231	10,6296 (1,1749-96,16768)
	T	Pele clara x não clara	0,0052	10,1584 (1,3134-78,5699)
	T/T	Pele clara x não clara	0,0033	10,6296 (1,3897-81,3040)
	C	Pele escura x clara	0,0223	11,0000 (1,2016-100,6921)
	T/C	Pele escura x clara	0,0231	10,6296 (1,1749-96,1676)
rs531268453¹ (69788550)	C	Pele média x clara	0,0182	8,66666 (1,0695-70,2287)
	T/C	Pele média x clara	0,0053	10,6296 (1,3516-83,5945)
	T	Cabelo loiro claro x não loiro claro	0,0383	78,3600 (3,0381-2021,0779)
	C/T	Cabelo loiro claro x não loiro claro	0,0383	76,7647 (3,0519-1930,8462)
rs114935806¹ (69788560)	T	Pele escura x não escura	0,0268	26,0091 (1,2315-549,2860)
	C/T	Pele escura x não escura	0,0270	25,6334 (1,2222-537,5900)
	C	Pele clara x escura	0,0214	18,7757 (0,9538-369,5688)
	C/C	Pele clara x escura	0,0218	18,3150 (0,9383-357,4809)
	G	Pele escura x não escura	0,0151	15,9622 (1,6292-156,3883)
	C/G	Pele escura x não escura	0,0154	15,5504 (1,6026-150,8869)
	G	Pele escura x clara	0,0214	18,7757 (0,9538-369,5688)
	C/G	Pele escura x clara	0,0218	18,3150 (0,9383-357,4809)
rs561386441¹ (69788574)	C	Cabelo loiro claro x não loiro claro	0,0382	78,6000 (3,0474-2027,2568)
	A/C	Cabelo loiro claro x não loiro claro	0,0382	77,0000 (3,0612-1936,7591)
	C	Cabelo preto x não preto	0,0042	15,8860 (1,8286-138,0081)
	T/C	Cabelo preto x não preto	0,0044	15,4294 (1,7895-133,0287)
	C	Olho castanho escuro x não castanho escuro	0,0063	17,6263 (0,9848-315,4754)
	T/C	Olho castanho escuro x não castanho escuro	0,0065	17,2652 (0,9686-307,7292)
	T	Pele clara x escura	0,0015	30,6504 (1,6655-564,0599)
	T/T	Pele clara x escura	0,0016	29,3162 (1,6073-534,6843)
	C	Pele escura x não escura	0,0005	27,3529 (3,1303-239,0099)
rs13322505² (69801975)	T/C	Pele escura x não escura	0,0006	26,1214 (3,0216-225,8180)
	C	Pele escura x média	0,0083	13,3333 (1,5208-116,8960)
	T/C	Pele escura x média	0,0088	12,7570 (1,4732-110,4673)

Tabela 11. Associações fenotípicas encontradas para 21 pontos de variação dentre os 133 analisados em toda a extensão do gene *MITF*. *Singletons* estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente. Associações significativas mesmo após aplicação da correção de Bonferroni estão em negrito, considerando $\alpha_{\text{corrigido}} = 0,0004$ (isto é, $0,05/133$ SNPs).

(continuação)

Ponto de variação	Alelo / Genótipo	Características	Valor de <i>p</i>	Odds ratio (95% IC)
rs13322505² (69801975)	C	Pele escura x clara	0,0015	30,6504 (1,6655-564,0599)
	T/C	Pele escura x clara	0,0016	29,3162 (1,6073-534,6843)
	T	Pele média x escura	0,0083	13,3333 (1,5208-116,8960)
	T/T	Pele média x escura	0,0088	12,7570 (1,4732-110,4673)
rs76992570² (69807750)	G/A	Pele escura	0,0261	26,5384 (1,2532-561,9627)
	A	Pele escura	0,0263	26,0810 (1,2419-547,6914)
rs80163893² (69812860)	T	Cabelo castanho claro x não castanho claro	0,0055	33,4444 (1,7045-656,1844)
	C/T	Cabelo castanho claro x não castanho claro	0,0056	32,7154 (1,6788-637,5374)
rs72950039² (69812903)	C	Pele clara x escura	0,0223	11,0000 (1,2016-100,6921)
	C/C	Pele clara x escura	0,0231	10,6296 (1,1749-96,1676)
	T	Pele escura x clara	0,0223	11,0000 (1,2016-100,6921)
	C/T	Pele escura x clara	0,0231	10,6296 (1,1749-96,1676)
rs2217456² (69887578)	A/A	Cabelo preto x loiro escuro	0,0037	12,3636 (1,5381-99,3810)
rs78124318² (69887807)	A	Cabelo loiro claro x preto	0,0154	12,6000 (1,8741-84,7105)
	G/A	Cabelo loiro claro x preto	0,0169	11,0869 (1,7583-69,9087)
	G	Cabelo preto x loiro claro	0,0154	12,6000 (1,8741-84,7105)
	G/G	Cabelo preto x loiro claro	0,0169	11,0869 (1,7583-69,9087)
rs112598669² (69985718)	A	Cabelo preto x não preto	0,0157	12,2926 (1,3546-111,5455)
	G/A	Cabelo preto x não preto	0,0161	12,0238 (1,3346-108,3251)
	G	Pele clara x escura	0,0057	24,6000 (1,3020-464,7693)
	G/G	Pele clara x escura	0,0060	23,7649 (1,2688-445,0988)
	A	Pele escura x não escura	0,0029	21,6923 (2,3768-197,9745)
	G/A	Pele escura x não escura	0,0031	20,9259 (2,3164-189,0367)
	A	Pele escura x média	0,0241	10,6923 (1,1678-97,8926)
	G/A	Pele escura x média	0,0249	10,3333 (1,1420-93,4952)
	A	Pele escura x clara	0,0057	24,6000 (1,3020-464,7693)
	G/A	Pele escura x clara	0,0060	23,7649 (1,2688-445,0988)
	G	Pele média x escura	0,0241	10,6923 (1,1678-97,8926)
	G/G	Pele média x escura	0,0249	10,3333 (1,1420-93,4952)
	G	Presença x ausência de sardas e pintas	0,0137	12,9113 (1,4223-117,2022)
	G/G	Presença x ausência de sardas e pintas	0,0141	12,6172 (1,4002-113,6914)
rs576634776³ (70008502)	T	Cabelo loiro claro x não loiro claro	0,0382	78,6000 (3,0474-2027,2568)
	C/T	Cabelo loiro claro x não loiro claro	0,0382	77,0000 (3,0612-1936,7591)
70014320³	T	Cabelo loiro claro x não loiro claro	0,0382	78,6000 (3,0474-2027,2568)
	C/T	Cabelo loiro claro x não loiro claro	0,0382	77,0000 (3,0612-1936,7591)
rs36118030³ (70014384)	G	Pele clara x escura	0,0211	18,9065 (0,9605-372,1326)
	G/G	Pele clara x escura	0,0215	18,4429 (0,9449-359,9707)
	A	Pele escura x não escura	0,0042	37,2242 (1,8953-731,0692)
	G/A	Pele escura x não escura	0,0043	36,3424 (1,8640-708,5533)
	A	Pele escura x média	0,0224	18,3831 (0,9338-361,8773)
	G/A	Pele escura x média	0,0228	17,9315 (0,9186-350,0112)

Tabela 11. Associações fenotípicas encontradas para 21 pontos de variação dentre os 133 analisados em toda a extensão do gene *MITF*. *Singletons* estão destacados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente. Associações significativas mesmo após aplicação da correção de Bonferroni estão em negrito, considerando $\alpha_{\text{corrigido}} = 0,0004$ (isto é, $0,05/133$ SNPs).

(conclusão)

Ponto de variação	Alelo / Genótipo	Características	Valor de <i>p</i>	<i>Odds ratio</i> (95% IC)
rs36118030³ (70014384)	A	Pele escura x clara	0,0211	18,9065 (0,9605-372,1326)
	G/A	Pele escura x clara	0,0215	18,4429 (0,9449-359,9707)
	G	Pele média x escura	0,0224	18,3831 (0,9338-361,8773)
	G/G	Pele média x escura	0,0228	17,9315 (0,9186-350,0112)
rs150314710⁴ (70015549)	C	Cabelo castanho claro x castanho escuro	0,0150	11,0714 (1,2639-96,9762)
	T/C	Cabelo castanho claro x castanho escuro	0,0158	10,6410 (1,2293-92,1077)
	T	Cabelo castanho escuro x castanho claro	0,0150	11,0714 (1,2639-96,9762)
	T/T	Cabelo castanho escuro x castanho claro	0,0158	10,6410 (1,2293-92,1077)
rs2131025⁴ (70015647)	T/C	Cabelo preto x ruivo	0,0014	10,2666 (2,6008-40,5261)
	T/T	Cabelo ruivo x preto	0,0009	11,7000 (2,9080-47,0729)
rs78240629⁴ (70016569)	A	Pele clara x escura	0,0211	18,9065 (0,9605-372,1326)
	A/A	Pele clara x escura	0,0215	18,4429 (0,9449-359,9707)
	G	Pele escura x não escura	0,0042	37,2242 (1,8953-731,0692)
	A/G	Pele escura x não escura	0,0043	36,3424 (1,8640-708,5533)
	G	Pele escura x média	0,0224	18,3831 (0,9338-361,8773)
	A/G	Pele escura x média	0,0228	17,9315 (0,9186-350,0112)
	G	Pele escura x clara	0,0211	18,9065 (0,9605-372,1326)
	A/G	Pele escura x clara	0,0215	18,4429 (0,9449-359,9707)
	A	Pele média x escura	0,0224	18,3831 (0,9338-361,8773)
	A/A	Pele média x escura	0,0228	17,9315 (0,9186-350,0112)
rs77962238⁴ (70016857)	A	Cabelo castanho claro x castanho escuro	0,0150	11,0714 (1,2639-96,9762)
	G/A	Cabelo castanho claro x castanho escuro	0,0158	10,6410 (1,2293-92,1077)
	G	Cabelo castanho escuro x castanho claro	0,0150	11,0714 (1,2639-96,9762)
	G/G	Cabelo castanho escuro x castanho claro	0,0158	10,6410 (1,2293-92,1077)

Dentre os haplótipos encontrados para as regiões individualizadas do gene *MITF* e para sua extensão completa, somente três haplótipos da região promotora, sete da região 5'UTR e CD, um na região 3'UTR e 20 na extensão total do gene foram associados a pelo menos um fenótipo de pigmentação, como mostrado nas Tabelas 12, 13, 14 e 15, respectivamente.

Tabela 12. Associações fenotípicas encontradas para três haplótipos dentre os 17 encontrados para a região promotora. Associações significativas mesmo após aplicação da correção de Bonferroni estão em negrito, considerando $\alpha_{\text{corrigido}} = 0,0029$ (isto é, $0,05/17$ haplótipos).

Haplótipo	Proporção Grupo A/Grupo B	Característica	Valor de <i>p</i>	<i>Odds Ratio</i> (IC95%)
Promo-1d1	0,0273/0,0018	Pele escura x não escura	0,0151	15,7289 (1,6207-152,6428)
	0,0454/0,0018	Pele escura x não escura	0,0006	26,7142 (3,0897-230,9779)
	0,0454/0,0000	Pele escura x clara	0,0015	29,8720 (1,6376-544,8841)
Promo-1h1	0,0454/0,0036	Pele escura x média	0,0081	13,0952 (1,5120-113,4114)
	0,0349/0,0000	Cabelo preto x não preto	0,0003	37,8288 (2,1196-675,1236)
	0,0349/0,0000	Cabelo preto x castanho escuro	0,0047	19,0900 (1,0682-341,1573)
Promo-1i1	0,0182/0,0000	Pele escura x não escura	0,0266	25,9216 (1,2358-543,6952)

Tabela 13. Associações fenotípicas encontradas para sete haplótipos dentre os 51 encontrados para a região 5'UTR e CD. Nenhuma associação permanece significativa após aplicação da correção de Bonferroni, considerando $\alpha_{\text{corrigido}} = 0,0010$ (isto é, $0,05/50$ haplótipos).

Haplótipo	Proporção Grupo A/Grupo B	Característica	Valor de <i>p</i>	<i>Odds Ratio</i> (IC95%)
5'CDS-1a1	0,0265/0,0000	Cabelo castanho claro x não castanho claro	0,0061	31,7058 (1,6260-618,2183)
5'CDS-1e2a	0,0187/0,0000	Pele escura x não escura	0,0291	24,5734 (1,1712-515,5516)
5'CDS-2c1	0,0342/0,0000	Olho verde x castanho escuro	0,0072	22,3215 (1,1920-417,9709)
5'CDS-2c2	0,0280/0,0000	Pele escura x não escura	0,0049	34,7320 (1,7807-677,4215)
5'CDS-2c21	0,0184/0,0000	Cabelo preto x não preto	0,0185	19,6479 (1,0093-382,4636)
5'CDS-2k1a1	0,0280/0,0019	Pele escura x não escura	0,0171	14,9134 (1,5362-144,7801)
5'CDS-3a	0,0333/0,0000	Olho azul x castanho escuro	0,0305	24,0598 (1,1401-507,7246)
	0,0155/0,0000	Olho castanho claro x não castanho claro	0,0423	19,4705 (0,9289-408,0938)

Tabela 14. Associações fenotípicas encontradas para um único haplótipo dentre os 18 encontrados para a região 3'UTR. Tal associação não permanece significativa após aplicação da correção de Bonferroni, considerando $\alpha_{\text{corrigido}} = 0,0028$ (isto é, 0,05/18 haplótipos).

Haplótipo	Proporção Grupo A/Grupo B	Característica	Valor de <i>p</i>	Odds Ratio (IC95%)
3UTR-1d1	0,0185/0,0000	Pele escura x não escura	0,0256	26,5023 (1,2634-555,9334)

Tabela 15. Associações fenotípicas encontradas para 20 haplótipos dentre os 132 encontrados para a extensão total do gene *MITF*. Nenhuma associação permanece significativa após aplicação da correção de Bonferroni, considerando $\alpha_{\text{corrigido}} = 0,0004$ (isto é, 0,05/132 haplótipos).

(continua)

Haplótipo	Frequência Grupo A/Grupo B	Característica	Valor de <i>p</i>	Odds Ratio (IC95%)
MITF-1	0,1100/0,0111	Pele clara x escura	0,0021	11,0000 (1,4590 - 82,9300)
	0,0691/0,0071	Cabelo castanho escuro x preto	0,0051	10,3257 (1,3344 - 79,8972)
MITF-2	0,1111/0,0071	Cabelo loiro claro x preto	0,0345	17,3750 (1,4909 - 202,4759)
	0,1304/0,0071	Cabelo loiro escuro x preto	0,0010	20,8500 (2,4383 - 178,2821)
	0,0850/0,0000	Pele clara x escura	0,0021	17,2615 (1,0264 - 290,2840)
MITF-16	0,0426/0,0000	Olho castanho claro x castanho escuro	0,0060	23,9171 (1,2749 - 448,6831)
MITF-19	0,0286/0,0029	Cabelo preto x não preto	0,0258	10,1470 (1,1240 - 91,6024)
MITF-24	0,0357/0,0000	Cabelo preto x não preto	0,0019	28,1291 (1,5448 - 512,1759)
	0,0444/0,0000	Pele escura x clara	0,0089	20,8612 (1,1110 - 391,6993)
	0,0444/0,0024	Pele escura x não escura	0,0043	18,9302 (2,0899 - 171,4647)
MITF-29	0,0333/0,0024	Pele escura x não escura	0,0199	14,0344 (1,4427 - 136,5237)
MITF-45	0,0333/0,0000	Pele escura x não escura	0,0057	32,6800 (1,6729 - 638,3785)
MITF-47	0,0667/0,0000	Olho mel x castanho escuro	0,0120	42,1929 (1,9761 - 900,8531)
	0,0667/0,0021	Olho mel x não mel	0,0101	33,3571 (2,9350 - 379,1110)
MITF-52	0,0222/0,0000	Pele escura x não escura	0,0324	23,0790 (1,0984 - 484,9153)
MITF-55	0,0263/0,0000	Cabelo castanho claro x não castanho claro	0,0242	27,5503 (1,3094 - 579,6319)
MITF-70	0,0222/0,0000	Pele escura x não escura	0,0324	23,0790 (1,0984 - 484,9153)
MITF-77	0,0213/0,0000	Olho castanho claro x não castanho claro	0,0353	21,8648 (1,0409 - 459,2685)
MITF-81	0,0556/0,0000	Cabelo ruivo x não ruivo	0,0370	80,3142 (3,1580 - 2042,5063)
MITF-82	0,0556/0,0000	Cabelo ruivo x não ruivo	0,0370	80,3142 (3,1580 - 2042,5063)
MITF-84	0,0556/0,0000	Cabelo ruivo x não ruivo	0,0370	80,3142 (3,1580 - 2042,5063)

Tabela 15. Associações fenotípicas encontradas para 20 haplótipos dentre os 132 encontrados para a extensão total do gene *MITF*. Nenhuma associação permanece significativa após aplicação da correção de Bonferroni, considerando $\alpha_{\text{corrigido}} = 0,0004$ (isto é, $0,05/132$ haplótipos).

(conclusão)

Haplótipo	Frequência Grupo A/Grupo B	Característica	Valor de p	Odds Ratio (IC95%)
MITF-97	0,0556/0,0000	Cabelo ruivo x não ruivo	0,0370	80,3142 (3,1580 - 2042,5063)
MITF-108	0,0556/0,0000	Cabelo loiro claro x não loiro claro	0,0370	80,3142 (3,1580 - 2042,5063)
MITF-114	0,0556/0,0000	Cabelo loiro claro x não loiro claro	0,0370	80,3142 (3,1580 - 2042,5063)
MITF-116	0,0556/0,0000	Cabelo ruivo x não ruivo	0,0370	80,3142 (3,1580 - 2042,5063)
MITF-122	0,0556/0,0000	Cabelo ruivo x não ruivo	0,0370	80,3142 (3,1580 - 2042,5063)

6.7. Interações TF/TFBS

A fim de entender as associações fenotípicas encontradas para variantes nas regiões regulatórias, foram estudadas as interações entre fatores de transcrição (*Transcription Factor*, TF) e sítios de ligação de fatores de transcrição (*Transcription Factor Binding Site*, TFBS) presentes nessas regiões. A Tabela 16 mostra os fatores de transcrição que podem interagir com as regiões onde se encontram as variantes associadas com algum fenótipo de pigmentação e a diferença da eficiência da ligação (Δ LOD) entre o fator e os alelos dessas variantes. Assim, valores positivos de Δ LOD indicam que o fator se liga mais eficientemente à sequência que contém o alelo alternativo.

Como é possível ver na Tabela 16, as seis variantes presentes na região promotora estão dentro de regiões hipersensíveis à DNase I e/ou regiões onde há acetilação de histonas (marca H3K27Ac) e valores positivos de Δ LOD foram encontrados principalmente para as variantes rs74845300 e rs80232752. Similarmente, variantes presentes em sítios regulatórios intrônicos coincidem com regiões hipersensíveis à DNase I e/ou regiões onde há acetilação de histonas. As variantes intrônicas rs13322505 e rs112598669 apresentaram variações positivas de LOD, e quatro das variantes intrônicas não estão inseridas em nenhuma região regulatória.

Deste modo, observa-se que o haplótipo promotor Promo-1h1 apresenta alelos que resultam em maior afinidade por fatores de transcrição, o que acarretaria em maior expressão do gene *MITF*, o que explica sua associação com fenótipos mais escuros.

Tabela 16. Fatores de Transcrição (TF) que podem ter sua eficiência de ligação (LOD) ao gene *MITF* alterada pelas variantes associadas à algum fenótipo de pigmentação presentes em regiões regulatórias (TFBS *Motif*).

(continua)

Ponto de Variação	REF	ALT	Posição	Região	TF	TFBS <i>Motifs</i>	Δ LOD
rs74845300	A	G	69786714 (-1.872)	Promotora Região hipersensível da DNase1	CEBPB	SIX5_disc2	12,0
						Znf143_disc1	11,9
rs80232752	A	T	69787573 (-1013)	Promotora Região hipersensível da DNase1 Marca H3K27Ac	STAT3, TBP e CTCF	CHOP::CEBPalpha	0,7
						STAT_known11	0,6
						Zbtb3	6,4
rs12489786	T	C	69788068 (-518)	Promotora Região hipersensível da DNase1 Marca H3K27Ac	STAT1	Foxj2_2	-1,0
						GR_known10	0,3
rs531268453	C	T	69788550 (-36)	Promotora Região hipersensível da DNase1	ZNF143, SIX5, POLR2A e MAS	Dados não encontrados	
rs114935806	C	T	69788560 (-26)	Promotora Região hipersensível da DNase1	ZNF143, SIX5, POLR2A, PHF8 e MAS	ATF3_known4	12,0
						BHLHE40_disc2	-1,4
						CAC-binding-protein	-11,9
						E2F_disc8	-8,6
						Ets_disc9	-2,7
						HEY1_disc2	-0,6
						Maf_known5	5,7
						Rad21_disc6	-6,4
						Rad21_disc7	-3,6
						STAT_disc7	0,2
						Sin3Ak-20_disc6	-0,8
						YY1_disc4	-1,1
YY1_disc5	-2,6						

Tabela 16. Fatores de Transcrição (TF) que podem ter sua eficiência de ligação (LOD) ao gene *MITF* alterada pelas variantes associadas à algum fenótipo de pigmentação presentes em regiões regulatórias (TFBS *Motif*).

(continuação)

Ponto de Variação	REF	ALT	Posição	Região	TF	TFBS <i>Motifs</i>	Δ LOD
rs561386441	C	G	69788574 (-12)	Promotora Região hipersensível da DNase1	ZNF143, SIX5, POLR2A, PHF8 e MAS	ZNF524	-
rs113085085	A	C	69801287 (+12.701)	Íntron	Dados não encontrados	AP-1_disc9	-11,9
						PAX-8_1	1,5
						PAX-8_2	1,0
rs13322505	T	C	69801975 (+13.389)	Íntron	Dados não encontrados	BCL_disc6	12,0
						Mrg_1	4,7
						NF-kappaB_known1	12,0
						PTF1-beta	2,3
						Tgif1_2	5,9
						Zic_2	1,1
rs76992570	G	A	69807750 (+19.164)	Íntron	Dados não encontrados	Dados não encontrados	
rs80163893	C	T	69812860 (+24.274)	Íntron Região hipersensível da DNase1 Marca H3K27Ac	NR3C1	Pax-5_disc3	-1,1
						p300_disc5	-1,9
rs72950039	C	T	69812903 (+24.317)	Íntron Região hipersensível da DNase1 Marca H3K27Ac	NR3C1	GR_known9	0,6
						Sox_16	2,1
rs2217456	A	G	69887578 (+98.992)	Íntron Região hipersensível da DNase1	JUND, POLR2A e p300	YY1_disc2	-12
rs78124318	G	A	69887807 (+99.221)	Íntron Marca H3K27Ac	Dados não encontrados	Cdc5	-11,9
						Pou2f2_known8	2,1
						Pou5f1_known2	4,3
						p300_disc5	-1,7

Tabela 16. Fatores de Transcrição (TF) que podem ter sua eficiência de ligação (LOD) ao gene *MITF* alterada pelas variantes associadas à algum fenótipo de pigmentação presentes em regiões regulatórias (TFBS *Motif*).

(conclusão)

Ponto de Variação	REF	ALT	Posição	Região	TF	TFBS <i>Motifs</i>	Δ LOD
rs112598669	G	A	69985718 (+197.132)	Íntron	Dados não encontrados	Mef2_known1	11,0
						Mef2_known4	2,3
rs150314710	T	C	70015549 (+226.963)	3'UTR Região de interação com miRNAs	Dados não encontrados	Mef2_known5	1,3
						CDP_2	-11,9
rs2131025	T	C	70015647 (+227.061)	3'UTR Região de interação com miRNAs	Dados não encontrados	CDP_6	-2,0
						GATA_known2	1,8
						GATA_known4	-1,0
						GATA_known9	2,5
						GR_disc6	1,1
rs78240629	A	G	70016569 (+227.983)	3'UTR Região de interação com miRNAs	Dados não encontrados	Mtf1_1	12,0
						Nr2f2	11,5
						Pax-4_5	1,8
						Pou2f2_disc1	-4,3
rs77962238	G	A	70016857 (+228.271)	3'UTR Região de interação com miRNAs	Dados não encontrados	CACD_2	-8,8
						Mef2_known1	-1,0
						Pax-4_4	-1,9
						Pou3f2_2	11,9
						Zic_2	-2,9

6.8. Interações miRNAs/mRNAs

Com o intuito de estudar as associações fenotípicas encontradas para variantes presentes na região 3'UTR, assim como de haplótipos completos dessa região, as interações miRNA/mRNA foram analisadas quanto a sua estabilidade (E) e especificidade (S). As Tabelas 17 e 18 trazem os miRNAs que apresentam maior variação de estabilidade e especificidade quando interagem com as diferentes sequências haplotípicas encontradas para a região 3'UTR do gene *MITF*. Adicionalmente, foram elencados na Tabela 19 os miRNAs cujas interações com os diferentes haplótipos da 3'UTR apresentam maior variação do valor de ES, que reflete tanto a estabilidade quanto a especificidade dessa interação.

Tabela 17. Valores de E para os 20 miRNAs com maior ΔE dentre os 150 mais estáveis. Valores de E menores que a média estão destacados em vermelho. Nas duas linhas inferiores estão apresentadas médias e desvios calculados para os 150 miRNAs mais estáveis.

miRNA	3UTR1	3UTR2	3UTR1a1	3UTR1a	3UTR1b	3UTR1d2a	3UTR2a	3UTR1c	3UTR1d2a1	3UTR1a1a	3UTR1d	3UTR1e	3UTR2b	3UTR1d1	3UTR1f	3UTR2c	3UTR1g	3UTR2d	ΔE
miR-6813-3p	0,8368	0,8068	0,0802	0,8020	0,8020	0,8068	0,8068	0,8020	0,7947	0,8020	0,8068	0,8020	0,8068	1,2730	0,8020	0,8068	0,8020	0,8020	1,1929
miR-6799-5p	0,7860	1,0127	0,0831	1,0129	1,0127	1,2236	0,7401	1,0129	1,2236	1,0129	1,0129	1,0129	1,0127	1,0129	1,0129	1,0127	1,0129	1,0127	1,1405
miR-6744-3p	0,9017	1,2097	0,0949	1,0004	1,0004	1,2238	0,7421	1,0004	1,2238	1,0004	1,0004	1,0004	1,2097	1,0004	1,0004	1,2108	1,0004	1,2097	1,1289
miR-7114-3p	0,7942	0,7524	0,0590	0,7524	0,7524	0,7524	0,4865	0,7524	0,7524	0,7524	0,7524	0,7524	0,7524	1,1770	0,7524	0,7524	0,7524	1,1305	1,1180
miR-6798-5p	0,7336	1,1956	0,0832	0,8508	0,8508	1,1832	0,7910	0,8508	1,1832	0,8508	0,8508	0,8508	1,1956	0,8508	0,8508	1,1955	0,8508	1,1956	1,1124
miR-6893-5p	0,8539	1,0856	0,0625	1,1084	1,0856	1,0856	0,7992	1,0856	1,0856	1,1084	1,0856	1,0856	1,0961	1,1460	1,0856	1,0856	1,0856	1,0856	1,0834
miR-7974	0,6456	0,7253	0,0477	0,7253	0,7253	0,7253	0,6686	0,7253	1,1262	0,7253	0,7253	0,7253	0,7253	0,7253	0,7253	0,7253	0,7253	0,7253	1,0786
miR-7110-5p	1,1103	1,1104	0,0598	1,1103	1,1102	1,1261	0,7657	1,1103	1,1261	1,1103	1,1103	1,1103	1,1104	1,1103	1,1103	1,1104	1,1103	1,1104	1,0663
miR-6810-3p	0,7871	0,9213	0,0809	0,9213	0,9213	0,9213	0,6711	0,9213	0,9213	0,9213	0,9213	0,9213	0,9213	1,1455	0,9213	0,9213	0,9213	0,9213	1,0647
miR-6873-3p	0,9240	1,1019	0,0672	1,0956	1,0956	1,1019	0,7562	1,0954	1,1252	1,0956	1,1019	1,0956	1,1019	1,0936	1,0956	1,1019	1,0956	1,0956	1,0580
miR-4747-3p	0,7220	0,8783	0,1941	0,8551	0,8510	0,8307	0,5805	0,8551	0,8307	0,8551	0,8551	0,8551	0,8783	1,2467	0,8551	0,8783	0,8551	0,8783	1,0526
miR-6776-3p	0,9179	1,1309	0,0876	1,0754	1,0713	1,0510	0,8449	1,0754	1,0510	1,0754	1,0754	1,0754	1,1309	1,0754	1,0754	1,1309	1,0754	1,1309	1,0433
miR-6865-3p	0,9229	1,1069	0,0690	1,1044	1,1044	1,1069	0,7835	1,1044	1,1096	1,1044	1,1069	1,1044	1,1069	1,1062	1,1044	1,1069	1,1044	1,1044	1,0406
miR-6851-5p	0,8269	1,0368	0,0719	1,0553	1,0407	1,1119	0,6603	1,0553	1,1119	1,0553	1,0553	1,0553	1,0368	1,0553	1,0553	1,0368	1,0553	1,0368	1,0400
miR-6870-5p	0,8308	1,0954	0,0678	1,0954	1,0954	1,0955	0,7944	1,0954	1,0955	1,0954	1,0954	1,0954	1,0954	1,0954	1,0954	1,0954	1,0954	1,0954	1,0278
miR-6788-5p	0,9204	1,1065	0,0851	1,0814	1,0814	1,1065	0,7764	1,0814	1,0988	1,0814	1,1065	1,0814	1,1065	1,1063	1,0814	1,1065	1,0814	1,0814	1,0213
miR-6819-3p	0,8955	1,0988	0,0788	1,0868	1,0868	1,0988	0,7746	1,0868	1,0946	1,0868	1,0988	1,0868	1,0988	1,0983	1,0868	1,0988	1,0868	1,0868	1,0201
miR-6769b-5p	0,6887	0,9109	0,0890	0,9109	0,9109	0,9242	0,7910	0,9217	0,9242	0,9217	0,9109	0,9109	1,1055	0,9109	0,9109	0,9109	0,9109	0,9109	1,0165
miR-4725-5p	0,6978	0,7413	0,1980	0,7413	0,7413	0,7413	0,4366	0,7413	0,7413	0,7413	0,7413	0,7413	0,7413	1,2048	0,7413	0,7413	0,7413	0,7413	1,0068
miR-6797-5p	0,7747	1,0746	0,0835	1,0541	1,0541	1,0746	0,8334	1,0541	1,0668	1,0541	1,0746	1,0541	1,0746	1,0754	1,0541	1,0746	1,0541	1,0541	0,9919
Média	1,0944	1,1501	0,8906	1,1476	1,1443	1,1570	0,8085	1,1434	1,1567	1,1480	1,1440	1,1430	1,1467	1,1487	1,1429	1,1530	1,1429	1,1521	
SD	0,2215	0,1683	0,5093	0,1706	0,1700	0,1744	0,1115	0,1710	0,1741	0,1698	0,1725	0,1717	0,1748	0,1644	0,1718	0,1663	0,1718	0,1643	

Tabela 18. Valores de S para os 20 miRNAs com maior ΔS dentre os 150 mais específicos. Valores de S menores que a média estão destacados em vermelho. Nas duas linhas inferiores estão apresentadas médias e desvios calculados para os 150 miRNAs mais específicos.

miRNA	3UTR1	3UTR2	3UTR1a1	3UTR1a	3UTR1b	3UTR1d2a	3UTR2a	3UTR1c	3UTR1d2a1	3UTR1a1a	3UTR1d	3UTR1e	3UTR2b	3UTR1d1	3UTR1f	3UTR2c	3UTR1g	3UTR2d	ΔS
miR-891a-5p	0,4577	2,9604	0,1206	2,9282	2,9200	2,8699	1,9655	2,9282	2,8699	2,9282	2,9282	2,9282	2,9604	2,9282	2,9282	2,9604	2,9282	2,9604	2,8398
miR-6732-3p	1,7671	2,3157	0,1164	2,3029	2,3029	2,3157	1,5367	2,3029	2,2742	2,3029	2,3157	2,3029	2,3157	2,7948	2,3029	2,3157	2,3029	2,3029	2,6784
miR-8072	0,3895	2,2825	0,0916	2,6060	2,6060	2,6427	1,6502	2,6060	2,6427	2,6060	2,6060	2,6060	2,6060	2,6060	2,6060	2,6060	2,6060	2,6060	2,5511
miR-365a-3p	0,9371	2,7503	0,2440	1,7831	1,7535	1,6181	1,8617	1,7831	1,6181	1,7831	1,7831	1,7831	2,7503	1,7831	1,7831	2,7503	0,8383	2,7503	2,5064
miR-6813-3p	0,6447	1,5600	0,1417	1,5472	1,5472	1,5600	1,5600	1,5472	1,5186	1,5472	1,5600	1,5472	1,5600	2,6460	1,5472	1,5600	1,5472	1,5472	2,5044
miR-4471	2,7167	2,2757	0,2912	2,0990	2,5959	2,0990	1,5087	2,0990	2,0990	2,0990	2,0990	2,0990	2,2757	2,0990	2,0990	2,2757	2,0990	2,2757	2,4255
miR-3688-5p	2,5671	2,1726	0,1725	2,1726	2,1726	2,1525	1,8335	2,1726	2,1525	2,1726	2,1726	2,1726	2,1726	2,1726	2,1726	2,1726	2,1726	2,1726	2,3946
miR-5582-3p	2,8367	1,9765	0,4437	1,9765	1,9765	2,7458	1,1356	2,1964	2,7458	2,1964	1,9765	1,9765	1,9765	1,9765	1,9765	1,9765	1,9765	1,9765	2,3930
miR-4275	1,1269	2,4375	0,0458	2,1366	2,1034	1,9491	1,6237	2,1366	1,9491	2,1366	2,1366	2,1366	2,4375	2,1366	2,1366	2,4375	2,1366	2,4375	2,3917
miR-3610	2,7210	2,7210	0,3538	1,8019	2,7210	2,7210	1,8000	2,7210	2,7210	2,7221	2,7210	2,7210	2,7164	2,7210	2,7210	2,7210	2,7210	2,7210	2,3683
miR-9-3p	1,7155	2,2825	0,1301	2,3555	2,3555	2,2825	1,4284	2,3555	2,4941	2,3555	2,2825	2,3555	2,2825	2,2402	2,3555	2,2825	2,3555	2,3555	2,3640
miR-3613-3p	0,9321	2,6381	0,2952	1,9549	1,9549	1,9549	1,9549	1,9549	1,9549	1,9549	1,9549	1,9549	2,6381	1,9549	1,9549	2,6381	1,9549	2,6381	2,3429
miR-365b-3p	1,3316	2,7503	0,4100	1,7831	1,7535	1,6181	1,8617	1,7831	1,6181	1,7831	1,7831	1,7831	2,7503	1,7831	1,7831	2,7503	1,7831	2,7503	2,3404
miR-944	1,5886	2,2842	0,1258	2,2842	2,2842	2,2842	1,7027	2,3823	2,3823	2,4596	2,3823	2,3823	2,2842	2,2842	2,3823	2,2842	2,3823	2,2842	2,3338
miR-4653-5p	2,3861	2,4127	0,0816	2,3861	2,3861	2,4127	1,5637	2,3861	2,3775	2,3861	2,4127	2,3861	2,4127	2,3571	2,3861	2,4127	2,3861	2,3861	2,3310
miR-19b-2-5p	0,7172	2,9025	0,5876	2,3367	2,3367	2,9025	1,9461	2,3367	2,8997	2,3367	2,9025	2,3367	2,9025	2,8799	2,3367	2,9025	2,3367	2,3367	2,3148
miR-3681-3p	0,8454	1,2906	0,1974	1,2906	1,2906	1,2906	1,0001	1,2906	1,2906	1,2906	1,2906	2,4864	1,2906	1,2906	1,2906	1,4602	1,2906	1,2906	2,2889
miR-5589-3p	0,6612	1,1113	0,3981	1,0681	1,0681	1,1113	1,1113	1,0681	1,0294	1,0681	1,1113	1,0681	1,1113	2,6818	1,0681	1,1113	1,0681	1,0681	2,2838
miR-4723-3p	2,3790	2,3834	0,1356	2,3790	2,3790	2,3834	1,6663	2,3790	2,3820	2,3790	2,3834	2,3790	2,3834	2,3834	2,3790	2,3834	2,3790	2,3790	2,2477
miR-548d-3p	1,3077	2,5119	0,2789	2,1582	2,1231	1,8831	1,5555	2,1582	1,8831	2,1582	2,1582	2,1582	2,5119	2,1582	2,1582	2,5119	2,1582	2,5119	2,2329
Média	2,2256	2,4853	1,9436	2,4469	2,4456	2,4537	1,6651	2,4468	2,4339	2,4576	2,4577	2,4465	2,4826	2,4888	2,4397	2,4869	2,4337	2,4686	
SD	0,7708	0,4212	1,1234	0,4508	0,4569	0,4464	0,3125	0,4538	0,4804	0,4462	0,4326	0,4373	0,4317	0,3870	0,4551	0,4162	0,4717	0,4433	

Tabela 19. Valores de *ES* para os 20 miRNAs com maior ΔES dentre os 150 de maior especificidade e estabilidade. Valores de *ES* menores que a média estão destacados em vermelho. Nas duas linhas inferiores estão apresentadas médias e desvios calculados para os 150 miRNAs com maior especificidade e estabilidade.

miRNA	3UTR1	3UTR2	3UTR1a1	3UTR1a	3UTR1b	3UTR1d2a	3UTR2a	3UTR1c	3UTR1d2a1	3UTR1a1a	3UTR1d	3UTR1e	3UTR2b	3UTR1d1	3UTR1f	3UTR2c	3UTR1g	3UTR2d	ΔES
miR-6813-3p	2,1104	2,9412	0,2885	2,9261	2,9261	2,9344	2,9412	2,9261	2,8717	2,9261	2,9412	2,9261	2,9412	4,8887	2,9261	2,9412	2,9261	2,9261	4,6003
miR-4638-5p	2,5881	4,2078	0,3852	4,0293	4,0293	4,0185	2,6716	4,0293	4,0185	4,0293	4,0241	4,0282	4,2078	4,2366	4,7971	4,2077	4,7971	4,2135	4,4119
miR-7114-3p	2,5777	2,5474	0,2770	2,5474	2,5474	2,5474	1,2462	2,5474	2,5474	2,5474	2,5474	2,5474	2,5474	4,6012	2,5474	2,5474	2,5474	2,5474	4,3242
miR-4725-5p	1,9843	2,3317	0,3737	2,3317	2,3317	2,3317	0,9880	2,3317	2,3317	2,3317	2,3317	2,3317	2,3317	4,6864	2,3317	2,3317	2,3317	2,3317	4,3127
miR-6773-3p	4,5798	3,9682	0,3335	3,9461	3,9461	3,9615	2,9024	3,9461	1,7812	3,9461	3,9682	3,9461	3,9682	3,8131	3,9461	3,9682	3,9461	3,9461	4,2464
miR-5589-3p	1,9098	2,1726	0,6485	2,1237	2,1237	2,1688	2,1726	2,1237	2,0662	2,1237	2,1726	2,1237	2,1726	4,6982	2,1237	2,1726	2,1237	2,1237	4,0497
miR-4747-3p	2,2750	2,1896	0,7286	2,0669	2,0470	1,9525	1,4220	2,0669	1,9525	2,0669	2,0637	2,0669	2,1896	4,7488	2,0669	2,1897	2,0669	2,1932	4,0201
miR-6744-3p	2,3535	4,2408	0,3777	3,6833	3,6833	4,3011	2,4085	3,6833	4,3011	3,6833	3,6833	3,6833	4,2408	3,6908	3,6833	4,2427	3,6833	4,2448	3,9233
miR-5090	2,2362	3,9430	0,3372	3,9913	3,9781	4,2598	2,3918	3,9913	4,2598	3,9913	3,9851	3,9913	3,9430	3,9987	3,9913	3,9556	3,9913	3,9492	3,9226
miR-6788-5p	3,2922	4,1734	0,2699	4,0207	4,0207	4,1658	2,8769	4,0207	4,1226	4,0207	4,1734	4,0207	4,1734	4,1863	4,0207	4,1734	4,0207	4,0207	3,9164
miR-6763-3p	2,9429	3,4506	0,3960	3,4170	3,4170	3,4456	2,1071	3,4170	3,3656	3,4170	3,4506	3,4170	3,4506	4,3015	3,4170	3,4506	3,4170	3,4170	3,9054
miR-6776-3p	3,1696	4,1800	0,3060	3,9739	3,9526	3,8319	3,3064	3,9739	3,8320	3,9739	3,9678	3,9739	4,1800	3,9790	3,9739	4,1800	3,9739	4,1871	3,8811
miR-7848-3p	2,6927	4,0950	0,2950	4,0917	4,0917	4,0890	2,9845	4,0917	4,1004	4,0917	4,0950	4,1327	4,0950	4,0988	4,1321	4,0950	4,1327	4,0917	3,8377
miR-3610	4,6698	4,6627	0,8718	3,4369	4,6698	4,6554	3,1218	4,6698	4,6554	3,4369	4,6627	4,6698	4,6480	4,6759	4,6698	4,6627	4,6698	4,6698	3,8041
miR-4669	2,2907	4,1496	0,3677	3,4085	3,2221	3,4085	2,2987	3,4085	3,4085	3,4085	3,4085	3,4085	4,1496	3,4161	3,4085	4,1496	3,4085	4,1543	3,7866
miR-501-3p	2,0719	3,9777	0,5910	3,9396	4,2923	3,9292	2,4329	3,9396	3,9292	3,9396	3,9355	3,9396	3,9777	3,9455	3,9396	3,9781	3,9396	3,9819	3,7013
miR-605-5p	4,1004	4,1310	0,4743	4,1004	4,1004	4,1254	2,7344	4,1004	2,9768	4,1004	4,1310	4,1004	4,1310	4,1165	4,1004	4,1310	4,1004	4,1004	3,6567
miR-548ah-5p	1,9111	4,0527	0,5263	4,1223	4,0581	4,0471	2,9013	4,0581	4,0471	4,1223	4,0527	4,0581	4,0924	4,0621	4,0581	4,0527	4,0581	4,0581	3,5960
miR-3171	2,6980	4,0289	0,7795	4,0316	4,0316	4,0260	2,6184	4,0316	4,0260	4,3673	4,0289	4,0316	4,0289	4,0361	4,0316	4,0289	4,0316	4,0316	3,5878
miR-6077	4,0993	4,1579	0,5923	4,0993	4,0993	4,1524	2,5428	4,0993	4,1280	4,0993	4,1579	4,0993	4,1579	4,1745	4,0993	4,1579	4,0993	4,0993	3,5821
Média	3,9914	4,2679	3,3531	4,2048	4,2049	4,2297	2,8639	4,2078	4,1715	4,2112	4,2269	4,2177	4,2606	4,2901	4,2061	4,2690	4,2176	4,2405	
SD	1,0232	0,7008	1,8924	0,7531	0,7569	0,7475	0,5343	0,7549	0,8238	0,7525	0,7281	0,7293	0,7196	0,6671	0,7576	0,6978	0,7490	0,7317	

É importante destacar que uma interação miRNA/mRNA específica pode não ser funcional se não for estável, e uma interação estável que seja pouco específica para a 3'UTR em questão pode ter pouca relevância na regulação da transcrição dos diferentes haplótipos encontrados para a 3'UTR do gene *MITF* que resultam nos diferentes fenótipos de pigmentação amostrados. Portanto, os miRNAs cujas interações sejam tanto estáveis quanto específicas para o gene *MITF*, podem estar relacionados com os diferentes fenótipos associados aos pontos de variação presentes na região 3'UTR. Adicionalmente, é importante enfatizar que caso tais miRNAs não sejam expressos nos tipos celulares envolvidos na pigmentação, eles não teriam como desempenhar seus papéis biológicos na regulação da expressão gênica pós-transcricional do *MITF*.

Dentre os 50 miRNAs apresentados nas tabelas acima, somente um (miR-6813-3p) se encontra entre os 20 selecionados pelos valores de ΔE , ΔS e ΔES , como mostrado na Tabela 20.

Tabela 20. Distribuição dos 50 miRNAs que apresentaram os maiores valores de ΔE , ΔS e ΔES nas interações com os diferentes haplótipos encontrados para 3'UTR do gene *MITF*.

ΔE	ΔS	ΔES	ΔE	ΔS	ΔES
	miR-19b-2-5p		miR-6744-3p		miR-6744-3p
		miR-3171			miR-6763-3p
	miR-3610	miR-3610	miR-6769b-5p		
	miR-3613-3p				miR-6773-3p
	miR-365a-3p		miR-6776-3p		miR-6776-3p
	miR-365b-3p		miR-6788-5p		miR-6788-5p
	miR-3681-3p		miR-6797-5p		
	miR-3688-5p		miR-6798-5p		
	miR-4275		miR-6799-5p		
	miR-4471		miR-6810-3p		
		miR-4638-5p	miR-6813-3p	miR-6813-3p	miR-6813-3p
	miR-4653-5p		miR-6819-3p		
		miR-4669	miR-6851-5p		
	miR-4723-3p		miR-6865-3p		
miR-4725-5p		miR-4725-5p	miR-6870-5p		
miR-4747-3p		miR-4747-3p	miR-6873-3p		
		miR-501-3p	miR-6893-5p		
		miR-5090	miR-7110-5p		
		miR-548ah-5p	miR-7114-3p		miR-7114-3p
	miR-548d-3p				miR-7848-3p
	miR-5582-3p		miR-7974		
	miR-5589-3p	miR-5589-3p		miR-8072	
		miR-605-5p		miR-891a-5p	
		miR-6077		miR-9-3p	
	miR-6732-3p			miR-944	

6.9. Relação entre índice de melanina e os fenótipos e genótipos

A fim de avaliar a diferença entre os fenótipos de pigmentação da pele estudados, foi analisada a variação do índice de melanina em áreas expostas e não expostas ao sol nos três grupos de cor de pele (clara, média e escura) (Figura 14), assim como nas diferentes classificações do IBGE (brancos, pardos, pretos e amarelos) (Figura 15).

Figura 14. Variação do índice de melanina nos indivíduos de pele clara (*box* branco), pele média (*box* cinza claro) e pele escura (*box* cinza escuro), em áreas expostas e não expostas ao sol.

Figura 15. Variação do índice de melanina dentro dos grupos de indivíduos declarados brancos (*box* branco), amarelos (*box* cinza claro), pardos (*box* cinza) e pretos (*box* cinza escuro), em áreas expostas e não expostas ao sol.

Como esperado, há um aumento do índice de melanina dos indivíduos de pele clara para os de pele escura, assim como dos indivíduos declarados brancos para os declarados pretos, sendo que os declarados amarelos apresentam um valor intermediário entre os brancos e os pardos. Há uma sobreposição dos valores de índice de melanina encontrados para indivíduos amarelos e brancos que resultou em uma diferenciação não significativa desses dois grupos quanto a esse parâmetro ($p > 0,05$). Adicionalmente, a diferenciação entre os grupos só não foi significativa para pardos e negros quanto aos valores obtidos para a área exposta, o que pode ser explicado pela grande representação de indivíduos com peles tipo III e IV entre os pardos e pela alta capacidade de bronzeamento destes (Figura 16), fazendo que com esses atinjam um tom de pele próximo aos declarados pretos. As figuras 14 e 15 também mostram um aumento claro do índice de melanina nas áreas expostas em relação as não expostas ao sol.

Figura 16. Distribuição da escala Fitzpatrick dentre os indivíduos auto classificados como brancos, pardos e pretos.

Como é possível ver na Tabela 21, a comparação dos valores de índice de melanina encontrados para indivíduos com diferentes genótipos nas 17 variantes que apresentaram associação com algum fenótipo de pigmentação (*singletons* não inclusos), mostrou que a distribuição de melanina entre os genótipos é significativa para 13 variantes (rs74845300, rs80232752, rs12489786, rs114935806, rs561386441, rs13322505, rs76992570, rs2217456, rs78124318, rs112598669, rs36118030, rs2131025, rs78240629) quando analisamos os valores para a área não exposta ao sol, e para 10 variantes (rs74845300, rs80232752, rs12489786, rs13322505, rs72950039, rs2217456, rs112598669, rs36118030, rs2131025, rs78240629), quando analisamos os valores obtidos para áreas expostas ao sol.

As diferenças entre os valores médios de índice de melanina deixaram de ser significativos nas áreas expostas ao sol para as variantes rs114935806, rs561386441, rs76992570, e rs78124318, o que pode indicar que os alelos alternativos dessas variantes tenham uma menor contribuição na determinação da cor da pele exposta ao sol. Entretanto, é importante ressaltar que os valores de probabilidade associado a elas são marginalmente significativos. Já para a variante rs72950039, os valores só se tornam significativamente diferentes nas áreas expostas ao sol, indicando que o alelo rs72950039*T pode conferir maior capacidade de bronzeamento à pele (Tabela 21).

Tabela 21. Comparação do índice de melanina considerando 17 pontos de variação associados a algum fenótipo de pigmentação (*singletons* não estão incluídos). Valores significativos de *p* (teste de Mann-Whitney e Kruskal-Wallis) estão indicados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(continua)

Ponto de Variação	REF (0)	ALT (1)	Área não exposta ao sol								Área exposta ao sol							
			Média (desvio padrão)			Mann-Whitney	Kruskal-Wallis	Diferenciação entre grupos <i>p</i>			Média (desvio padrão)			Mann-Whitney	Kruskal-Wallis	Diferenciação entre grupos <i>p</i>		
			0/0	0/1	1/1	<i>p</i>	<i>p</i>	0/0 x 0/1	0/0 x 1/1	0/1 x 1/1	0/0	0/1	1/1	<i>p</i>	<i>p</i>	0/0 x 0/1	0/0 x 1/1	0/1 x 1/1
rs74845300 ¹ (69786714)	A	G	1,1066 (0,4322)	1,6597 (0,4122)	-	0,0054	-	-	-	-	1,5328 (0,3805)	1,9112 (0,2371)	-	0,0146	-	-	-	-
rs80232752 ¹ (69787573)	A	T	1,1066 (0,4322)	1,6597 (0,4122)	-	0,0054	-	-	-	-	1,5328 (0,3805)	1,9112 (0,2371)	-	0,0146	-	-	-	-
rs12489786 ¹ (69788068)	T	C	1,1038 (0,4304)	1,4280 (0,5104)	-	0,0166	-	-	-	-	1,5286 (0,3755)	1,8110 (0,4409)	-	0,0101	-	-	-	-
rs114935806 ¹ (69788560)	C	T	1,1114 (0,4322)	2,1175 (0,0116)	-	0,0224	-	-	-	-	1,5366 (0,3810)	2,0577 (0,1919)	-	0,0689	-	-	-	-
rs561386441 ¹ (69788574)	C	G	1,1102 (0,4331)	1,7141 (0,4827)	-	0,0239	-	-	-	-	1,5345 (0,3795)	1,9711 (0,4006)	-	0,0551	-	-	-	-
rs13322505 ² (69801975)	T	C	1,1009 (0,4271)	1,9191 (0,2403)	-	0,0004	-	-	-	-	1,5227 (0,3718)	2,2540 (0,1623)	-	0,0001	-	-	-	-
rs76992570 ² (69807750)	G	A	1,1142 (0,4467)	1,9900 (0,0297)	-	0,0370	-	-	-	-	1,5341 (0,3876)	2,0462 (0,1822)	-	0,0802	-	-	-	-
rs80163893 ² (69812860)	C	T	1,1174 (0,4392)	1,0446 (0,2099)	-	0,9464	-	-	-	-	1,5414 (0,3828)	1,3652 (0,1335)	-	0,4195	-	-	-	-
rs72950039 ² (69812903)	C	T	1,1110 (0,4368)	1,3197 (0,4432)	-	0,1494	-	-	-	-	1,5309 (0,3784)	1,8519 (0,3826)	-	0,0149	-	-	-	-
rs2217456 ² (69887578)	A	G	1,3326 (0,4343)	1,0926 (0,4589)	1,0291 (0,3950)	-	0,0001	<0,001	<0,001	>0,05	1,7104 (0,3685)	1,5135 (0,3499)	1,4795 (0,3639)	-	0,0008	<0,01	<0,01	>0,05
rs78124318 ² (69887807)	G	A	1,1388 (0,4415)	0,8706 (0,3013)	-	0,0010	-	-	-	-	1,5530 (0,3872)	1,3921 (0,2770)	-	0,0520	-	-	-	-
rs112598669 ² (69985718)	G	A	1,1092 (0,4335)	1,6575 (0,4263)	-	0,0130	-	-	-	-	1,5335 (0,3813)	1,9458 (0,1613)	-	0,0140	-	-	-	-
rs36118030 ³ (70014384)	G	A	1,1088 (0,4318)	1,9698 (0,0408)	-	0,0097	-	-	-	-	1,5355 (0,3805)	1,9992 (0,1523)	-	0,0393	-	-	-	-

Tabela 21. Comparação do índice de melanina considerando 17 pontos de variação associados a algum fenótipo de pigmentação (*singletons* não estão incluídos). Valores significativos de *p* (teste de Mann-Whitney e Kruskal-Wallis) estão indicados em vermelho. Os números 1, 2, 3 e 4 associados aos pontos de variação indicam se estes estão presentes nas regiões promotora, íntron, éxon e 3'UTR, respectivamente.

(conclusão)																		
Ponto de Variação	REF (0)	ALT (1)	Área não exposta ao sol								Área exposta ao sol							
			Média (desvio padrão)			Mann-Whitney	Kruskal-Wallis	Diferenciação entre grupos <i>p</i>			Média (desvio padrão)			Mann-Whitney	Kruskal-Wallis	Diferenciação entre grupos <i>p</i>		
			0/0	0/1	1/1	<i>p</i>	<i>p</i>	0/0 x 0/1	0/0 x 1/1	0/1 x 1/1	0/0	0/1	1/1	<i>P</i>	<i>p</i>	0/0 x 0/1	0/0 x 1/1	0/1 x 1/1
rs150314710 ⁴ (70015549)	T	C	1,1154 (0,4351)	1,1503 (0,5183)	-	0,9271	-	-	-	-	1,5398 (0,3743)	1,5396 (0,5627)	-	0,9025	-	-	-	-
rs2131025 ⁴ (70015647)	T	C	1,1459 (0,4394)	0,9310 (0,3730)	1,7227 (0,0240)	-	0,0005	<0,01	>0,05	<0,05	1,5625 (0,3825)	1,4011 (0,3514)	1,8678 (0,0917)	-	0,0085	<0,05	>0,05	>0,05
rs78240629 ⁴ (70016569)	A	G	1,1104 (0,4315)	1,9698 (0,0408)	-	0,0097	-	-	-	-	1,5364 (0,3808)	1,9992 (0,1523)	-	0,0396	-	-	-	-
rs77962238 ⁴ (70016857)	G	A	1,1154 (0,4351)	1,1503 (0,5183)	-	0,9271	-	-	-	-	1,5398 (0,3743)	1,5396 (0,5627)	-	0,9025	-	-	-	-

7. Discussão

7. Discussão

7.1. Amostra populacional

A população brasileira é uma das mais miscigenadas, resultado da mistura entre europeus, africanos e ameríndios, principalmente (Pena et al. 2011), gerando uma população com características fenotípicas de pigmentação que vão dos tons mais claros aos mais escuros. Segundo censo do IBGE de 2010, na região de Ribeirão Preto 69,8% dos indivíduos se declararam brancos, 6,4% pretos, 22,8% pardos, 0,9% amarelos e 0,1% indígenas. A população amostrada aleatoriamente nesse estudo apresentou proporções semelhantes às do censo, sendo que 65,5% dos indivíduos se declararam brancos, 8,3% se declararam pretos, 24,3% se declararam pardos, 1,4% se declararam amarelos e 0,5% se declararam indígenas.

Contudo, essa mistura não é tão uniforme em todo o território brasileiro. Diversos estudos mostraram que indivíduos no norte (dados do Amazonas, Pará e Amapá) apresentam um componente ameríndio mais acentuado, chegando a 52%, e baixa contribuição africana (Guerreiro et al. 1993), indivíduos no nordeste (dados do Maranhão, Bahia, Ceará, Pernambuco e Alagoas) apresentam contribuição africana mais acentuada, chegando a 49% (Felix et al. 2010). De maneira geral a ancestralidade europeia é predominante em todas as regiões do país, e atinge valores mais altos no sul e sudeste, chegando a uma contribuição de 86% (Rodrigues de Moura et al. 2015). Isso mostra que estudos envolvendo ancestralidade ou fenótipos que possam depender dela devem ser feitos regionalmente.

Outro ponto levantado por alguns estudos é que as categorias de cor autodeclaradas utilizadas pelo IBGE (branco, pardo, preto, amarelo e indígena), assim como a ancestralidade podem não refletir corretamente a cor da pele dos indivíduos em diferentes regiões do país. Pena et al. (2011) explica que a percepção de cor, no Brasil, vai além do tom de pele e depende também de fatores como o formato do rosto e o tipo de cabelo. Além disso, as diferentes regiões do país recebem diferentes incidências solares e, com isso, indivíduos no norte e nordeste do país costumam ter pele mais escura e indivíduos no sul e sudeste costumam ter pele mais clara. Como consequência, indivíduos auto classificados como pretos ou pardos nas diferentes regiões do país possuem contribuições ancestrais (europeia, africana e ameríndia) e tons de pele diferentes. Isso fica evidente quando se observa que indivíduos autodeclarados pardos no Norte

(Pará, área de alta incidência solar) possuíam, em média, ancestralidade 68,6% europeia, 20,9% ameríndia e 10,6% africana, enquanto que indivíduos auto declarados pardos no Sul (Rio Grande do Sul, área de baixa incidência solar) possuíam, em média, ancestralidade 44,2% europeia, 11,4% ameríndia e 44,4% africana (Pena et al. 2011).

Como mostrado na Figura 15, os indivíduos declarados pardos na população estudada, possuem valores de índice de melanina na área exposta ao sol que se sobrepõem aos valores de índice de melanina das áreas não expostas de indivíduos declarados pretos. Assim, é possível que, tomando somente o tom de pele da área exposta ao sol, como cor do rosto e braços, indivíduos que possuem naturalmente (quando não exposta ao sol) pele moderadamente pigmentada possam se auto declarar como pretos. O mesmo ocorre com indivíduos declarados brancos, que, quando expõem a pele ao sol, e a depender da capacidade de bronzeamento, podem ter tons de pele semelhantes aos indivíduos declarados pardos. Como mostra a Figura 16, há uma grande variação na escala Fitzpatrick dentro das categorias branco, preto e pardo, realçando que a capacidade de se bronzear apresentada pelas peles tipo 3 a 5 podem interferir na percepção de cor que leva à classificação desses indivíduos como brancos, pardos e pretos. Além disso, na população estudada, não é possível distinguir brancos e amarelos considerando somente a cor da pele. Assim, como dito anteriormente (Pena et al. 2011), , não só o tom de pele, mas também características como o formato dos olhos, nariz e lábios assim como o tipo de cabelo, devem ter sido consideradas na classificação em branco, pardo, preto e amarelo.

7.2. *Variant Calling*

O *calling* das variantes do gene *MITF* foi feito por meio de dois algoritmos distintos, *HaplotypeCaller* e *UnifiedGenotyper*, ambos componentes do pacote GATK. Enquanto a ferramenta *UnifiedGenotyper* faz os *callings* analisando as sequências base a base, ou seja, verificando somente a posição em questão, o *HaplotypeCaller* utiliza uma abordagem diferente, em que avalia o que ele chama de ‘regiões ativas’, que são regiões que podem conter variantes em uma ou mais posições. Dessa maneira, o *HaplotypeCaller* é capaz de identificar *indels* com mais competência, o que é uma grande vantagem. Porém, essa ferramenta possui filtros mais restritivos e acaba não identificando variantes raras e *singletons* que costumam ser encontrados pelo *UnifiedGenotyper* (Van der Auwera et al. 2013).

Em concordância, nesse estudo, na região do gene *MITF*, a ferramenta *HaplotypeCaller* encontrou 10 *indels* não identificados pelo *UnifiedGenotyper*, que por sua vez identificou sozinho 4 variantes, duas delas *singletons*, que não foram encontradas pelo *HaplotypeCaller*.

7.3. Desvios do Equilíbrio de Hardy-Weinberg

Populações suficientemente grandes em que os casamentos ocorram ao acaso e que não estejam sob a influência de forças seletivas, como mutações, migrações e seleção natural, tendem a manter as frequências alélicas e genotípicas de uma geração para outra. Desvios significativos dessas proporções podem indicar que o locus em questão está sob alguma força seletiva, ou até, erros de genotipagem (Hartl e Clark 2007).

Dentre os 133 pontos de variação avaliados, rs146187921 e rs2217456 desviaram significativamente do equilíbrio de Hardy-Weinberg ($p = 0,0039 \pm 0,0001$ e $p = 0,0112 \pm 0,0001$, respectivamente). A variante rs146187921 apresentou um excesso de heterozigotos (aproximadamente 20%) que pode ser causado por um erro de amostragem, se os indivíduos coletados não representarem corretamente a população estudada. Isso ocorre mais frequentemente quando número de indivíduos amostrado é pequeno. Outra possibilidade é que haja uma força seletiva que favoreça o genótipo heterozigoto (seleção balanceadora) (Waples 2015). No entanto, essa variante não foi associada a nenhum fenótipo estudado nesse trabalho, e uma pressão seletiva atuando sobre uma determinada variante deixaria assinaturas nos sítios vizinhos.

Em contraste, a variante rs2217456 apresentou uma deficiência de heterozigotos (aproximadamente 19%) que pode ter sido causada por erros de genotipagem, caso heterozigotos tenham sido erroneamente genotipados como homozigotos. Isso ocorre mais frequentemente em casos em que um dos alelos é raro (Graffelman et al. 2015). Contudo, os alelos da variante em questão estão bem representados na amostra (A:0,4555 e G:0,5445). A existência de estruturação na população amostrada (efeito Wahlund) eventualmente é responsável pela deficiência de heterozigotos observada em um locus (Hartl e Clark 2007), o que estaria de acordo com o fato de ter sido encontrada uma associação fenotípica para essa variante (genótipo A/A associado a cabelos pretos). Entretanto, efeito Wahlund também afetaria sítios vizinhos, levando a um maior número de desvios. Desvios do Equilíbrio de Hardy-Weinberg observados para a variante rs2217456 dentro dos grupos fenotípicos (pele média,

pele escura, ausência de sardas e pintas, olho castanho escuro e cabelo preto), especialmente nas amostragens não aleatórias, são esperados, já que essa separação em grupos fenotípicos não obedece às premissas desse equilíbrio.

Deste modo, conclui-se que estes dois desvios podem ser interpretados como achados casuais, ou então representar erros de genotipagem. Como dito anteriormente, a distribuição dos haplótipos, tanto para as regiões promotora, 5'UTR e CD e 3'UTR do gene *MITF*, quanto para sua extensão completa, está de acordo com o esperado segundo o equilíbrio de Hardy-Weinberg.

7.4. Associações fenotípicas

Grande parte das variantes associadas a fenótipos complexos, como a pigmentação, não afetam diretamente as proteínas, mas sim os mecanismos de expressão dessas proteínas (Ward e Kellis 2016).

Elementos como fatores de transcrição e regiões regulatórias como *enhancers* e *insulators* são alguns elementos presentes nas proximidades da região promotora dos genes e que estão envolvidos na regulação da expressão gênica, facilitando ou inibindo a descompactação da cromatina e o início da transcrição (Danko et al. 2015). Dados obtidos pelo Haploreg e RegulomeDB, que trazem um conjunto de informações geradas pelo ENCODE (*Encyclopedia of DNA Elements*) Consortium, foram consultados a fim de encontrar elementos regulatórios atuantes nas regiões alteradas pela presença de variantes da promotora que eventualmente levem a uma alteração na expressão do gene *MITF*. Apesar da importância dos fatores de transcrição no sistema regulatório transcricional, poucos tiveram sua atividade verificada experimentalmente (Peled et al. 2016).

Elementos regulatórios também podem estar presentes em regiões intrônicas, regulando o *splicing* que levará a expressão de diferentes isoformas proteicas (Wang et al. 2012). Essas regiões podem ser encontradas nos bancos Haploreg e RegulomeDB, e também por meio do software SnpEff. Esse software também foi utilizado para estudar o efeito causado por variantes presentes na região codificante do gene *MITF*. Variações não sinônimas, isto é, que causam mudança de aminoácido, levam a alterações na estrutura proteica podendo alterar a função dessa proteína, enquanto que as variações sinônimas, aquelas que não causam mudança de

aminoácido, podem afetar a conformação, *splicing* e a estabilidade do pré-mRNA, eventualmente afetando os níveis de expressão da proteína (Teng et al. 2012).

Já as variações presentes na região 3'UTR do gene *MITF* podem afetar a interação miRNA/mRNA. miRNAs são pequenas sequências de RNA que atuam na regulação pós transcricional podendo inibir a tradução do mRNA em proteína (Bartel 2004). Dessa forma, haplótipos da 3'UTR podem ser regulados diferentemente por miRNAs, afetando a expressão do gene *MITF*. A fim de estudar as associações encontradas para variantes presentes na região 3'UTR, foram avaliados os miRNAs que interagem diferentemente com os 18 haplótipos encontrados para essa região do gene *MITF*, através do software mirhunt (Porto et al. 2015).

Apesar do papel regulatório do gene *MITF* na melanogênese, que envolve a regulação de genes como a tirosinase, *TYRP1* e *DCT*, cujas proteínas atuam diretamente na conversão de tirosina em melanina, e *OCA2* e *BCL2*, que também atuam nessa via, ainda não foram encontradas associações fortes entre variantes nesse gene (*MITF*) e fenótipos de pigmentação. A variante rs149617956 foi associada a olhos não azuis em um estudo que procurava diferentes suscetibilidades ao desenvolvimento de melanoma em indivíduos com fenótipos claros e escuros. Contudo a associação encontrada foi fraca ($OR = 2,01$, $p = 0,018$) (Yokoyama et al. 2011).

O presente estudo encontrou associações com fenótipos de pigmentação para 21 pontos de variação do gene *MITF*, além de três haplótipos da região promotora, sete haplótipos na região 5'UTR e CD, um haplótipo na região 3'UTR e 20 haplótipos do gene completo. A variante rs149617956 não foi associada a nenhum fenótipo de pigmentação e foi encontrada em somente um indivíduo (pele clara, olhos castanhos claros e cabelo preto).

7.4.1. Promotora

Seis variantes da região promotora foram associadas a fenótipos de pigmentação. Uma dessas variantes, rs531268453, é um *singleton* e o alelo T e o genótipo C/T foram associados a cabelos loiro claros ($OR = 78,3600$, $p = 0,0383$ e $OR = 76,7447$, $p = 0,0383$). É provável que essa associação tenha sido significativa somente pelo fato de que o número de indivíduos que compõem o grupo de loiros claros amostrados foi baixo ($n = 13$), configurando uma frequência de quase 8% nesse grupo. A frequência do alelo T nas populações do projeto 1000 genomas (<http://www.internationalgenome.org>) não mostra nenhum indicativo de que esse alelo

realmente esteja associado a fenótipos mais claros, já que não é encontrado na população europeia (Tabela 22), onde fenótipos claros são mais frequentes (Park et al. 2012; Lock-Andersen et al. 1998).

As variantes rs74845300 e rs80232752 tiveram seus alelos alternativos associados a fenótipos escuros, principalmente cabelos pretos ($OR = 41,0993$ $p = 0,002$). Nas populações mundiais esses alelos só são encontrados em populações africanas e americanas, o que é consistente com as associações encontradas. Essas duas variantes estão em desequilíbrio de ligação perfeito (Figura 10) inclusive nas populações africanas e americanas estudadas pelo projeto 1000 genomas. Além disso, como mostrado na Tabela 16, essas variantes se encontram em regiões regulatórias (região hipersensível da DNase1, e marca H3K27Ac) onde os TF CEBPB (rs74845300), STAT3, TBP e CTCF (rs80232752) podem atuar regulando a transcrição do gene *MITF*. Os TFBS alterados por essas variantes geram interações mais efetivas ($\Delta LOD > 0$) quando o alelo alternativo está presente. Sendo assim, é possível que haja uma interação mais intensa entre TF e TFBS resultante da presença do alelo alternativo dessas variantes, causando uma expressão intensificada do fator MITF e, conseqüentemente, aumentando a produção de eumelanina.

O fator CEBPB (CCAAT/*enhancer binding protein*) é um fator de transcrição que atua como um homodímero ou em associação com proteínas relacionadas (CEBP-alpha, CEBP-delta e CEBP-gamma). Kaikkonen et al. demonstrou, através de ensaios ChIP-Seq, que esse fator atua ativando a transcrição gênica ao se ligar a regiões regulatórias promotoras de seus genes alvo. Essa interação foi associada a um aumento da acetilação da histona H3, o que caracteriza *enhancers* ativos. Adicionalmente, foi observado que regiões caracterizadas pela metilação da histona H3, são enriquecidas com sítios reconhecidos por fatores como CEBPB, que mostrou se ligar, também, a novos *enhancers* (Kaikkonen et al. 2013). Adicionalmente, CEBPB pode atuar desmetilando histonas (H3K9me3), promovendo a expressão (Guo et al. 2015).

STAT3 (*signal transducer and activator of transcription 3*) é um ativador de transcrição que atua como homodímero ou heterodímero em associação com outros fatores da mesma família. Esses dímeros se ligam ao DNA em regiões ricas em AT, importantes na regulação da expressão gênica e organização da cromatina (Timofeeva et al. 2012). A mutação da variante rs80232752 (A -> T) pode contribuir para o aumento da afinidade desse fator pela promotora do gene *MITF*.

TATA-binding protein (TBP) é uma subunidade do fator de transcrição IID (TFIID), cuja ligação ao DNA (TATA-box) é o primeiro passo para a formação do complexo de pré-iniciação da transcrição pela RNA polimerase II. Após o recrutamento do fator se dá início ao

desenovelamento da cromatina, que é facilitado pela composição rica em AT (Boeva 2016). O alelo rs80232752*T pode contribuir para que essa região seja mais ativa.

O fator CTCF (*CCCTC-binding factor*) é membro de uma família de fatores com 11 sítios de zinco que podem interagir com diferentes sequências do DNA dependendo de sua composição metálica. Esse fator pode atuar em conjunto com acetil-transferases na acetilação de histonas, funcionando como ativador da transcrição, ou em conjunto com desacetilases, atuando como repressor da transcrição. Ainda, esse fator interage preferencialmente com DNA não metilado modulando o acesso a cromatina (Merkenschlager e Nora 2016).

Portanto, como dito anteriormente, a presença dos alelos rs74845300*G e rs80232752*T pode aumentar a efetividade da ligação ($\Delta\text{LOD} > 0$) dos fatores CEBPB, STAT3, TBP e CTCF à região promotora do gene *MITF*, intensificando sua expressão e, consequentemente, a produção de eumelanina. Os valores de índice de melanina (Tabela 21) corroboram essa hipótese, já que são significativamente diferentes para as duas variantes tanto nas regiões não expostas ao sol, quanto nas expostas ($p = 0,0054$ e $p = 0,0146$, respectivamente). Além disso, o haplótipo Promo-1h1, que contém rs74845300*G e rs80232752*T, foi significativamente associado à pele escura e cabelo preto ($\text{OR} = 26,7142$, $p = 0,0006$ e $\text{OR} = 37,8288$, $p = 0,0003$, respectivamente). Essas associações se mantiveram significativas mesmo após a correção de Bonferroni ($\alpha = 0,0029$).

A variante rs12489786 teve seu alelo alternativo associado a tons de pele mais escuros. Contudo, esse alelo possui frequência mais acentuada em populações sul asiáticas (aproximadamente 10%) e leste asiáticas (aproximadamente 8%), enquanto que nas outras populações a frequência está em torno de 2%. É possível que a variante rs12489786*C possa contribuir para fenótipos mais pigmentados em indivíduos asiáticos, principalmente no sul da Ásia, que possui indivíduos com tons de pele mais escuros (Day et al. 2016). Dentre os indivíduos amostrados nesse estudo, os que apresentam o alelo C nessa posição foram classificados como brancos, pardos ou pretos e possuem descendência, na maior parte, europeia e brasileira. Além disso, apesar dos valores médios de índice de melanina (Tabela 21) para os dois alelos serem significativamente diferentes ($p = 0,0166$ e $p = 0,0101$, para áreas não expostas e expostas ao sol respectivamente), as modificações causadas a um dos dois TFBS encontrados levam a uma interação menos efetiva ($\Delta\text{LOD} = -1$) enquanto que para o outro sítio o aumento é muito baixo ($\Delta\text{LOD} = 0,3$). O fator de transcrição STAT1 (*signal transducer and activator of transcription 1*), indicado como um possível regulador para essa região, assim como STAT3, interage com regiões ricas em AT, podendo, então, ter afinidade reduzida pela promotora do gene *MITF* quando o alelo C está presente. Em resumo, não há explicação

plausível para associação da variante rs12489786*C com fenótipos de pigmentação mais escuros na população estudada, e o fato de essa associação não ser significativa após a conservadora correção de Bonferroni, pode indicar um falso positivo.

As duas outras variantes da região promotora, rs114935806 e rs561386441, também apresentaram associação de seus alelos alternativos com fenótipos escuros. Como é possível ver na Tabela 22, esses alelos estão presentes com frequências polimórficas (> 1%) somente em populações africanas, o que é consistente com a associação encontrada. Porém, os *motifs* alterados pelo alelo T da variante rs114935806, de maneira geral, resultam em uma ligação menos efetiva dos FT, com exceção dos *motifs* ATF3_known4, Maf_known5 e STAT_disc7 ($\Delta\text{LOD} = 12$, $\Delta\text{LOD} = 5,7$ e $\Delta\text{LOD} = 0,2$, respectivamente), enquanto que o alelo G da variante rs561386441 altera o *motif* ZNF524 (valores de LOD não encontrados). Segundo dados obtidos pelo Haploreg, os fatores de transcrição ZNF143, SIX5, POLR2A e MAZ podem interagir com ambas regiões e o fator PHF8 pode interagir com a região que compreende rs561386441.

O fator ZNF143 (*zinc finger protein 143*) é um fator que pode atuar como regulador da transcrição alterando a compactação da cromatina, muitas vezes em conjunto com CTCF (Bailey et al. 2015). SIX5 (*SIX homeobox 5*), que atua na formação da retina (Winchester et al. 1999), também pode atuar em conjunto com ZNF143 na regulação gênica (Foley e Sidow 2013). POLR2A (*polymerase (RNA) II (DNA directed) polypeptide A*) é a maior subunidade da RNA polimerase II e auxilia na interação da polimerase com a fita de DNA (Choudhuri e Carlson 2008). MAZ (*Myc-associated zinc finger protein*) é um fator que se liga a regiões ricas em GC, sendo um importante fator de regulação da transcrição de genes que não possuem a região TATA-box e, na regulação do gene p53, atua como supressor (Lee et al. 2016). O fator PHF8 (*PHD Finger Protein 8*), que pode interagir com a região que contém a variante rs561386441, possui atividade de desmetilase das histonas 3 e 4 (H3K9me2, H3K9me1, H3K27me2 e H4K20me1), podendo atuar como ativador da transcrição de seus genes alvo (Shao et al. 2016).

É possível que as alterações causadas pelas duas variantes nos *motifs* encontrados nessa região, especialmente ATF3_known4, Maf_known5 e STAT_disc7, possam intensificar a interação de alguns dos fatores citados acima, causando uma transcrição mais intensa do gene *MITF* que levaria a um fenótipo mais escuro. Os valores de índice de melanina (Tabela 21) encontrados para heterozigotos portadores dos alelos rs114935806*T e rs561386441*G são significativamente mais altos que os encontrados para homozigotos para alelos referência para áreas não expostas ao sol ($p = 0,0224$ e $p = 0,0239$, respectivamente), porém essa diferença deixa de ser significativa quando olhamos valores referentes a áreas expostas ao sol ($p = 0,0689$).

e $p = 0,0551$, respectivamente). Isso pode indicar que a contribuição desses alelos para o aparecimento do fenótipo escuro seja baixa, e outros fatores como a exposição ao sol tenham um efeito mais acentuado sobre a cor da pele.

Além da Promo-1h1 citada acima, os haplótipos Promo-1d1 e Promo-1i1 também foram associados a pele escura (OR = 15,7289, $p = 0,0151$ e OR = 25,9216, $p = 0,0266$, respectivamente). Esses haplótipos contêm as variantes rs12489786*C e rs561386441*G (Promo-1d1) e rs114935806*T (Promo-1i1) cujas associações com fenótipos de pigmentação mais escuros já foram justificadas acima. As associações encontradas para os haplótipos Promo-1h1, 1d1 e 1i1 com fenótipos mais escuros de pigmentação reafirmam as hipóteses apresentadas de que as variantes rs74845300 e rs80232752, rs12489786 em conjunto com rs561386441, e rs114935806 podem estar contribuindo para o desenvolvimento de fenótipos mais escuros na população estudada.

Tabela 22. Frequências alélicas das variantes da região promotora do gene *MITF* nas populações mundiais.

Ponto de Variação		AFR	AMR	EUR	EAS	SAS
rs74845300	A	0,977	0,997	1,000	1,000	1,000
	G	0,023	0,003	-	-	-
rs80232752	A	0,977	0,997	1,000	1,000	1,000
	T	0,023	0,003	-	-	-
rs12489786	T	0,980	0,973	0,983	0,919	0,896
	C	0,020	0,027	0,017	0,081	0,104
rs531268453	C	0,998	0,997	1,000	1,000	1,000
	T	0,002	0,003	-	-	-
rs114935806	C	0,964	1,000	1,000	1,000	0,999
	T	0,036	-	-	-	0,001
rs561386441	C	0,984	0,999	0,999	1,000	0,999
	G	0,016	0,001	0,001	-	0,001

7.4.2. 5'UTR e CD

Na região CD foram encontradas associações para 8 variantes intrônicas e 3 exônicas. As 3 variantes exônicas estão presentes nos éxons 9 (rs576634776) e 10 (chr3:70014320 e rs36118030). rs576634776*T e chr3:70014320*T correspondem a *singletons* e foram associadas, novamente, a cabelo loiro claro. Porém, nas populações mundiais o alelo rs576634776*T é encontrados em frequência extremamente baixa somente na população

africana enquanto que a outra variante não foi reportada nessas populações (Tabela 23). O software SnpEff não forneceu dados de efeitos para essas variantes, pois não estão inclusas no pacote dbSNP utilizado (Build 138). Contudo, utilizando o software IGV foi possível verificar que a mutação rs576634776 (C -> T) é sinônima e, portanto, não afetaria a função do fator transcrito. É possível que, assim como para a variante rs531268453, a associação encontrada para rs576634776 seja um falso positivo e tenha sido significativa por conta da pequena amostragem de indivíduos de cabelo loiro nesse estudo, já que não há informações na literatura que corroborem essa associação. Já a mutação na posição chr3:70014320 (C -> T) causa uma mudança de aminoácido (Arg -> Lys). Essa mudança na estrutura do transcrito poderia resultar no mal funcionamento ou inatividade do fator MITF, resultando em fenótipos mais claros, como o cabelo loiro, por exemplo.

A terceira variante exônica (rs36118030) teve seu alelo alternativo associado a tons mais escuros de pele. Nas populações mundiais o alelo rs36118030*A é encontrado somente nas populações africana e americana, em concordância com as associações encontradas nesse estudo. Dados obtidos pelo software SnpEff mostram que essa mutação é sinônima (não causa mudança de aminoácido), sendo classificada como de baixo impacto. Como dito anteriormente, variações sinônimas podem alterar os níveis de expressão gênica se causarem alguma alteração na conformação e estabilidade do pré-mRNA. Um estudo mostrou que mutações sinônimas podem, inclusive, intensificar a tradução do transcrito em proteína (Gotea et al. 2015). Assim, é possível que a alteração causada por rs36118030 leve à uma alteração na estabilidade do transcrito que resulte na tradução intensificada do fator MITF, levando a uma maior produção de eumelanina, uma vez que os valores de índice de melanina obtidos para os heterozigotos em comparação com os homozigotos GG são significativamente diferentes tanto para a área não exposta quanto para a exposta ao sol ($p = 0,0097$ e $p = 0,0393$, respectivamente). Além disso, em concordância com o que foi hipotetizado acima, o haplótipo 5'CD-1e2a que contém rs36118030*A também foi associado a pele escura (OR = 24,5734, $p = 0,0291$).

O número reduzido de variantes encontradas em regiões exônicas indica que esse gene possui uma sequência conservada. Esses resultados estão em concordância com o proposto por McEvoy et al., de que o gene *MITF* é conservado em humanos e provavelmente é um dos fatores responsáveis pelo escurecimento da pele após a perda da pelagem característica de primatas e que provia proteção à radiação solar (McEvoy et al. 2006).

Dentre as 8 variantes intrônicas, uma é *singleton*. Como ocorreu para os outros *singletons* descritos, a variante rs113085085*C está associada a cabelos loiros. No entanto, nesse caso o alelo foi encontrado somente na população europeia (também *singleton*) em

concordância com a associação encontrada. Contudo, a frequência do alelo associado é muito baixa e não pode ser responsável pelo aparecimento do fenótipo na população europeia, já que essa população possui uma predominância de fenótipos mais claros (Lock-Andersen et al. 1998). Portanto, assim como para os outros *singletons*, essa associação pode ser um falso positivo.

As sete variantes intrônicas restantes não aparentam ter efeito na regulação da transcrição do gene *MITF* segundo resultados obtidos pelo SnpEff. Porém, como mostrado na Tabela 16, quatro dessas variantes estão presentes em regiões regulatórias como regiões hipersensíveis à DNaseI (rs80163893, rs72950039 e rs2217456) e sítios de acetilação de histonas (H3K27Ac) (rs80163893, rs72950039 e rs78124318).

Estudos mostram que o *splicing* alternativo ocorre simultaneamente à transcrição e, por isso, acredita-se que elementos que regulem a compactação da cromatina, como acetilação de histonas (H3K27Ac) podem também estar envolvidos na regulação de *splicing* (Matveeva et al. 2016). Sendo assim, as três variantes presentes em regiões onde ocorre acetilação de histonas, podem alterar a eficiência desse processo, diminuindo ou intensificando a expressão gênica.

A variante rs80163893*T foi associada a cabelo castanho claro (Tabela 11), contudo os valores de índice de melanina para os genótipos heterozigoto e homozigoto CC não são significativamente diferentes tanto para áreas expostas como para áreas não expostas ao sol. Sendo assim, somando o fato de não ser significativa após a correção de Bonferroni, é provável que a associação dessa variante, assim como do haplótipo 5'CD-1a1 que a contém, com cabelo castanho claro seja um falso positivo.

A variante rs72950039 teve seu alelo C associado a pele clara e o alelo T associado a pele escura. Em concordância, o alelo T só foi encontrado em populações africanas e americanas (Tabela 23). Além disso, os valores de índice de melanina obtidos para os dois genótipos (C/C e C/T) são significativamente diferentes, porém, somente para a área exposta ao sol ($p = 0,0149$). Isso pode ser um indicativo de que essa variante pode estar relacionada com a capacidade de bronzear. Um estudo verificou que o fator NR3C1 (*Nuclear Receptor Subfamily 3 Group C Member 1*), fator de transcrição indicado pelo Haploreg (Tabela 16), está presente em concentrações maiores na pele após irradiação com UVB (Chen e Zhang 2016). Esse fator pode estar atuando de maneira a intensificar a expressão de *MITF* quando a pele é exposta ao sol, intensificando a produção de pigmento, ou seja, promovendo o bronzeamento da pele.

A terceira variante intrônica presente em região regulatória, rs78124318, teve seu alelo A associado a cabelo loiro claro ($p = 0,0154$), enquanto que o alelo G foi associado a cabelo

preto ($p = 0,0154$). Nas populações mundiais, o alelo A está presente em maior frequência nas populações americanas, europeias e sul asiáticas. Apesar dos índices de melanina encontrados para os genótipos G/G e G/A serem significativamente diferentes, mesmo que somente para a área não exposta ao sol ($p = 0,0010$), em um primeiro momento parece pouco provável que esse alelo (rs78124318*A) realmente contribua para o aparecimento do fenótipo em questão, em vista que a população com maior quantidade de indivíduos loiros possui frequência desse alelo menor que de populações americanas e sul asiáticas que, em sua maioria, possuem indivíduos de cabelos mais pigmentados (Lock-Andersen et al. 1998; Park et al. 2012; Adhikari et al. 2016). Por outro lado, interações epistáticas podem ser responsáveis por efeitos específicos em determinadas populações, como já observado para outros genes relacionados à pigmentação (Branicki et al. 2009). Este raciocínio também é válido para o haplótipo 5'CD-3a associado a olho azul (OR = 24,0598 $p = 0,0305$), que contém a variante rs78124318*A.

Quanto ao restante das variantes intrônicas associadas a algum fenótipo de pigmentação, rs13322505*C foi associada a fenótipos escuros de pele, olhos e cabelo na população estudada e é encontrada somente em populações africanas, além de ter valores de índice de melanina para o genótipo T/C significativamente mais altos que os obtidos para o homozigoto T/T tanto para áreas não expostas ao sol, quanto para as expostas ($p = 0,0004$ e $p = 0,0001$, respectivamente); rs76992570*A foi associado a pele escura e é encontrada com frequência de pelo menos 1% somente em populações africanas; além disso, os valores de índice de melanina para o genótipo G/A são significativamente mais altos que os obtidos para o genótipo G/G em áreas não expostas ao sol ($p = 0,0370$), mas não para áreas expostas ($p = 0,0802$). A variante rs2217456 teve o genótipo A/A associado a cabelo preto, o que é condizente com o fato do alelo A ser menos frequente na população europeia, e com o fato de os valores de índice de melanina terem sido significativamente diferentes entre os genótipos A/A, A/G e G/G, exceto quando comparamos valores G/G e A/G, o que é compatível com efeito recessivo, indicando que a presença de apenas uma cópia do alelo A não deve ser suficiente para o aparecimento do fenótipo mais escuro. Finalmente, rs112598669*A foi associado a pele escura e cabelo preto (Tabela 11), condizente com a presença desse alelo exclusivamente em populações africanas e americanas (Tabela 23), enquanto que o alelo G foi associado à presença de sardas. Os valores de índice de melanina encontrados para os genótipos G/G e G/A se diferenciam significativamente tanto em áreas não expostas quanto para áreas expostas ao sol ($p = 0,0130$ e $p = 0,0140$, respectivamente). Apesar da confluência entre os dados acima, não foram encontradas informações suficientes na literatura para justificar as associações apresentadas.

Semelhantemente, as associações encontradas para os haplótipos 5'CD-2c2 e 5'CD-2c2a (rs13322505*C e rs2217456*G), 5'CD-1e2a (rs76992570*A e rs36118030*A), 5'CD-2k1a1 (rs2217456*G) e 5'CD-2c1 (rs112598669*A) não puderam ser justificadas até o momento.

Tabela 23. Frequências alélicas das variantes da região codificante do gene *MITF* nas populações mundiais.

Ponto de		AFR	AMR	EUR	EAS	SAS
rs113085085	A	1,000	1,000	0,999	1,000	1,000
	C	-	-	0,001	-	-
rs13322505	C	0,957	1,000	1,000	1,000	1,000
	T	0,043	-	-	-	-
rs76992570	G	0,971	0,996	1,000	1,000	1,000
	A	0,029	0,004	-	-	-
rs80163893	C	1,000	0,993	0,999	1,000	1,000
	T	-	0,007	0,001	-	-
rs72950039	C	0,914	0,990	1,000	1,000	1,000
	T	0,086	0,010	-	-	-
rs2217456	G	0,274	0,337	0,627	0,218	0,287
	A	0,726	0,663	0,373	0,782	0,713
rs78124318	G	0,998	0,952	0,969	0,992	0,969
	A	0,002	0,048	0,031	0,008	0,031
rs112598669	G	0,968	0,997	1,000	1,000	1,000
	A	0,032	0,003	-	-	-
rs576634776	C	0,999	1,000	1,000	1,000	1,000
	T	0,001	-	-	-	-
chr3:70014320	C	-	-	-	-	-
	T	-	-	-	-	-
rs36118030	G	0,976	0,996	1,000	1,000	1,000
	A	0,024	0,004	-	-	-

7.4.3. 3'UTR

Quatro variantes e um haplótipo da 3'UTR foram associados a fenótipos de pigmentação. Como dito anteriormente, alterações nessa região podem interferir na regulação pós-transcricional feita pela interação miRNA/mRNA, levando a diferentes níveis de expressão do fator MITF.

A variante rs150314710 teve o alelo C, assim como o genótipo T/C associados a cabelo castanho claro (em comparação com cabelo castanho escuro), e o alelo T, assim como o genótipo T/T associados a cabelo castanho escuro (em comparação com cabelo castanho claro) (Tabela 11). Nas populações mundiais, o alelo C tem baixas frequências em todas as

populações, não ultrapassando 5% em nenhuma delas (Tabela 24). Além disso, os valores de índice de melanina obtidos para os genótipos T/T e T/C não diferem de maneira significativa. Apesar dessa medida ter sido feita na pele, ela pode refletir a capacidade da presença dos diferentes genótipos de gerar fenótipos mais escuros ou mais claros de olhos e cabelos, podendo ser um indicativo de que (a) existem miRNAs específicos expressos em melanócitos relacionados ao desenvolvimento dos fios de cabelo, ou (b) a associação em questão possa ser um falso positivo.

A variante rs2131025 teve seus genótipos T/T e T/C associados a cabelo ruivo e cabelo preto, respectivamente, quando comparados um frente ao outro (Tabela 11). Nas populações globais, o alelo C é encontrado com maior frequência nas populações europeia e asiáticas, em desacordo com as associações encontradas, uma vez que as populações europeias possuem uma maior quantidade de indivíduos ruivos, fenótipo mais raro no restante dos continentes (Lock-Andersen et al. 1998; Adhikari et al. 2016).

A variante rs78240629*G foi associada a pele de tons mais escuros enquanto que o alelo A está associado a tons de pele mais claros. Em concordância com as associações encontradas, o alelo G é encontrado com frequências acima de 1% somente em populações africanas. Além disso, os valores de índice de melanina encontrados para os genótipos A/A e A/G diferem significativamente tanto nas áreas não expostas ao sol quanto nas expostas ($p = 0,0097$ e $p = 0,0396$). O haplótipo 3'UTR-1d1, que contém a variante rs78240629*G, também foi associado a pele escura (OR = 26,5023, $p = 0,0256$).

Finalmente, a variante rs77962238*A foi associada a cabelo castanho claro (quando comparado com cabelo castanho escuro) e o alelo G foi associado a cabelo castanho escuro (quando comparado com cabelo castanho claro). Nas populações globais, o alelo A só não é encontrado com frequência acima de 1% nas populações do leste da Ásia, o que não é suficiente para descartar essa associação ou dar embasamento a ela, já que indivíduos com cabelos castanhos são encontrados em todas as populações (Lock-Andersen et al. 1998; Adhikari et al. 2016; Park et al. 2012). Porém, caso a variante realmente contribua para a obtenção desses fenótipos, seria esperado que o alelo A, associado ao fenótipo mais claro, fosse mais frequente na população europeia. A Tabela 24 mostra o contrário, o alelo A tem frequências pouco maiores na África e no sul da Ásia. Além disso, valores de índice de melanina obtidos para os genótipos G/G e G/A não diferem significativamente e essa associação deixou de ser significativa após a correção de Bonferroni, dois indicativos de que essa associação pode ser um falso positivo. Contudo, a pigmentação é uma característica poligênica, e a análise dessa variante isoladamente pode não refletir completamente seu papel na formação do fenótipo.

Tabela 24. Frequências alélicas das variantes da região 3'UTR do gene *MITF* nas populações mundiais.

Ponto de variação		AFR	AMR	EUR	EAS	SAS
rs150314710	T	0,972	0,984	0,989	0,994	0,969
	C	0,028	0,016	0,011	0,006	0,031
rs2131025	T	0,996	0,955	0,912	0,965	0,934
	C	0,004	0,045	0,088	0,035	0,066
rs78240629	A	0,972	0,996	0,999	1,000	1,000
	G	0,028	0,004	0,001	-	-
rs77962238	G	0,974	0,984	0,990	0,994	0,969
	A	0,026	0,016	0,010	0,006	0,031

7.4.3.1. Interações miRNA/mRNA

Como dito anteriormente, foram analisadas as interações miRNA/mRNA para os diferentes haplótipos da 3'UTR do gene *MITF* a fim de encontrar miRNAs que possam atuar na regulação pós-transcricional do gene, resultando nos diferentes fenótipos encontrados. Foram analisadas três pontuações: *E*, que representa a estabilidade da interação; *S*, que representa a especificidade dessa interação; e *ES*, um valor que compreende tanto a especificidade quanto a estabilidade da interação miRNA/mRNA.

Então, analisando a Tabela 17, que traz os valores referentes a estabilidade, em relação ao haplótipo 3'UTR-1d1 que foi associado a pele escura, vemos que os miRNAs miR-6813-3p, miR-7114-3p, miR-4747-3p e miR-4725-5p podem interagir de maneira mais intensa com essa UTR em comparação com as outras. Analisando a Tabela 18, que traz valores de especificidade, vemos que os miRNAs miR-6732-3p, miR-6813-3p e miR-5589-3p podem se ligar à UTR em questão de maneira mais específica que a outras UTRs. Finalmente, a Tabela 19 traz os valores de *ES* que refletem tanto a especificidade quanto a estabilidade da interação entre o miRNA e o haplótipo da 3'UTR em questão, portanto, quanto maior o valor de *ES* mais específico é a interação de um determinado miRNA com o haplótipo da 3'UTR e mais estável será essa interação e, portanto, menor a expressão do fator MITF, uma vez que a regulação pós-transcricional mediada pelos miRNAs envolve a inibição da tradução do mRNA. O maior valor médio de *ES* foi encontrado justamente para a interação entre o haplótipo 3'UTR-1d1 (4,2690) e o miR-6813-3p. Os miRNAs miR-7114-3p, miR-4725-5p, miR-5589-3p, miR-4747-3p e miR-6763-3p também apresentaram valores altos de *ES* para a interação com essa UTR. Esses achados vão de encontro com a associação da 3'UTR-1d1 com fenótipo de pele escura, que dependeria de uma expressão mais intensa do fator MITF. Contudo, independente da

estabilidade e especificidade desses miRNAs com a referida UTR, tais miRNAs só poderão exercer papel biológico se forem expressos nos tecidos em estudo (pele, olhos e bulbo capilar). Uma busca rápida no *Human miRNA Expression Database (miRmine)*, <http://guanlab.ccmb.med.umich.edu/mirmine/index.html>) encontrou dados de expressão desses miRNAs no bulbo capilar, mas não na pele.

A regulação pós-transcricional do gene *MITF* já foi estudada em alguns trabalhos que mostraram o envolvimento de miRNAs (miR-137, miR-218, miR-182, miR-25 e miR-155) nesse processo (Bemis et al. 2008; Guo et al. 2014; Segura et al. 2009; Arts et al. 2015; Shi et al. 2016), porém esses miRNAs não estão entre os que possuem os maiores valores de ΔE , ΔS e ΔES encontrados nesse estudo. Joyce et al. identificou que os miRNAs miR-501-3p, miR-548d-3p e miR-944 são diferencialmente expressos em pele saudável e afetada por psoríase. Os miRNAs citados tiveram expressão aumentada nos tecidos afetados pela doença cujos sintomas envolvem a despigmentação da pele (Joyce et al. 2011). Os valores de especificidade e estabilidade encontrados nesse estudo referentes à 3'UTR-1d1 para esses miRNAs estão abaixo da média para essa UTR, podendo indicar que a menor especificidade e estabilidade desses miRNAs com a UTR em questão permite a expressão mais intensa do fator MITF e, conseqüentemente, o desenvolvimento de fenótipos mais pigmentados. Contudo, os valores de especificidade e estabilidade encontrados para esses miRNAs não são os mais baixos dentre as UTRs.

7.4.4. Singletons

A associação de *singletons* a fenótipos de pigmentação não era esperada pois, por terem frequências populacionais muito baixas, essas variantes não podem ser responsáveis pelo aparecimento de um determinado fenótipo em uma população. Como explicado anteriormente, essas associações se devem as frequências infladas desses alelos dentro de um grupo de poucos indivíduos, como os de cabelo loiro ($n = 13$). Contudo, a quantidade de *singletons* dentro desse grupo chamou a atenção.

Examinando com mais atenção a distribuição dos *singletons* encontrados dentro dos grupos fenotípicos de olhos e cabelos, esperaríamos que a distribuição de singletons dentro dos grupos ocorresse de maneira casual, sendo que em grupos com maior número de indivíduos encontraríamos um maior número de *singletons*. Porém, por alguma razão, fenótipos menos

pigmentados, como olho azul e cabelo loiro claro apresentam maior proporção de *singletons* (Figura 17). Contudo, essa diferença não é estatisticamente significativa (p do Teste Exato de Fisher $> 0,05$) para nenhuma das categorias fenotípicas, indicando que a flutuação mostrada na Figura 17 ocorreu ao acaso.

Figura 17. Proporção de indivíduos que possuem *singletons* (linha) dentre o total de indivíduos (colunas) dentro de cada categoria de cor de cabelo (a) e olhos (b).

7.4.5. Associações encontradas para a extensão completa do gene *MITF*

Foram encontradas 20 associações fenotípicas com os haplótipos completos do gene *MITF* (Tabela 15). Dessas 20, 13 (MITF-1, 16, 19, 24, 29, 45, 47, 70, 84, 97, 108, 114 e 116) têm em sua sequência variantes que foram associadas a algum fenótipo de pigmentação (Tabela 11) e cujas associações foram discutidas anteriormente.

MITF-1 foi associado a pele clara, o que está de acordo com a presença dos alelos rs74845300*A, rs80232752*A, rs12489786*T e rs561386441*C que também foram associados a fenótipos mais claros. Já os haplótipos MITF-19, 24, 29 e 70 foram associados a fenótipos de pigmentação mais escuros como cabelo preto (MITF-19 e 24) e pele escura (MITF-24, 29 e 70) e contém o alelo rs2217456*A cujo genótipo homozigoto A/A foi associado a cabelo preto. Além disso, o haplótipo MITF-24, possui ainda os alelos rs74845300*G e rs80232752*T que também foram associados a cabelo preto e pele escura e olhos castanhos escuros. Como discutido no item 5.1, esses fenótipos podem ser resultado interação diferenciada entre elementos regulatórios da região promotora do gene *MITF* e fatores de transcrição devido a presença de um ou outro alelo.

O haplótipo MITF-16 foi associado a olho castanho claro, quando comparado com olho castanho escuro. O alelo rs2131025*C presente nesse haplótipo, é menos frequente em populações africanas (< 1%) (Tabela 24), em concordância com a associação encontrada. A Tabela 21 mostra que os valores de índice de melanina não diferenciam significativamente os genótipos T/T e C/C ($p = >0,05$). Além disso, rs2131025 está em forte desequilíbrio de ligação com rs576 e rs704246, apresentando o haplótipo CCC que possui frequência de 8% na população amostrada (Figura 12). Os alelos rs576*C e rs704246*C são mais frequentes na população europeia (0,743 e 0,755, respectivamente), o que também está de acordo com o fenótipo encontrado. Olhando a Tabela 19 vemos que o haplótipo 3'UTR-1a1, que compõe MITF-16, apresenta o menor valor médio de ES, indicando que, frente as outras UTRs, esse haplótipo interage de maneira pouco específica e fraca com os miRNAs selecionados. Isso levaria ao desenvolvimento de um fenótipo mais escuro, discordando da associação encontrada.

Os haplótipos MITF-47, 84, 97, 108, 114 e 116 foram associados a fenótipos claros como olho mel (MITF-47), cabelo loiro claro (MITF-108 e 114) e cabelo ruivo (MITF-84, 97 e 116). Esses haplótipos contém os alelos rs2217456*A (MITF-97, 108, 116 e 47), rs150314710*C (MITF-114), rs2131025*C (MITF-84 e 116) e rs77962238*A (MITF-114), cujas associações, segundo o que foi discutido anteriormente, não estão de acordo com as distribuições alélicas dentro das populações mundiais e os fenótipos associados a esses haplótipos. Associações com fenótipos raros como olho mel ($n = 20$) e cabelos loiros claros ($n = 13$) e ruivos ($n = 11$) podem ter se mostrado significativas por conta de a amostragem desses fenótipos ter sido baixa e, por conta disso, alelos raros tiveram suas frequências alélicas infladas nesses grupos. Portanto, é possível que essas associações sejam falso positivos. O mesmo raciocínio se aplica aos haplótipos MITF-81, 82 e 122, associados a cabelo ruivo. Contudo, vale ressaltar que a pigmentação é uma característica poligênica e, por tanto, interações mais complexas, como epistasia, poderiam explicar as associações encontradas (Branicki et al. 2009).

O haplótipo MITF-45 foi associado a pele escura e em sua sequência está presente o alelo rs561386441*G que também foi associado a pele escura, o que explicaria a associação haplótipo/fenótipo encontrada. Porém, como discutido anteriormente (Item 5.1), a contribuição desse alelo para o desenvolvimento do fenótipo escuro não é grande, uma vez que após a exposição ao sol, a diferença entre os valores médios de índice de melanina obtidos para os genótipos homozigoto e heterozigoto, deixam de ser significativamente diferentes (Tabela 21).

Os haplótipos MITF-2, 52, 55 e 77 se diferenciam do restante em pontos de variação que não possuem associações fortes ($OR < 10$) a fenótipos de pigmentação e, por isso, ainda não foram discutidos aqui.

MITF-2, associado a cabelos não pretos e pele clara, possui a variante rs62253052*A que está em forte desequilíbrio de ligação com as variantes rs9837720*G, rs1427421*G e rs7651891*G (Figura 10). Como mostra a Tabela 25, esses alelos possuem frequências altas em populações europeias, em acordo com as associações encontradas. Foram encontradas associações fracas ($OR < 10$) desses alelos com fenótipos de pigmentação mais claros: rs62253052*A e cabelo loiro claro ($OR = 3,1958$, $p = 0,0314$); rs9837720*G e cabelos loiro escuro ($OR = 4,0959$, $p = 0,0163$); rs1427421*G e olhos azuis ($OR = 2,7918$, $p = 0,0481$); rs7651891*G e cabelo castanho escuro (quando comparado com cabelo preto, $OR = 2,6364$, $p = 0,0012$). Além dessas variantes, as rs74845300*A e rs80232752*A, associadas a pele clara e cabelo castanho escuro, também compõem esse haplótipo.

Assim como MITF-2, o haplótipo MITF-55, que foi associado a cabelo castanho claro, também apresenta a composição rs62253052*A, rs9837720*G, rs1427421*G e rs7651891*G, mas se diferencia do haplótipo anterior pelas variantes rs6761994*G, rs4855323*T e rs62250968*C, que estão em forte desequilíbrio de ligação na região codificante (Figura 11) e foram associadas a cabelos loiros ($OR = 3,8947$ e $p = 0,0000$, $OR = 2,8431$ e $p = 0,0285$, $OR = 2,9032$ e $p = 0,0314$, respectivamente). A Tabela 25 mostra que as frequências desses alelos são mais baixas em populações europeias (onde se espera ter maior ocorrência de fenótipos claros). No entanto, fenótipos intermediários como olhos e cabelos castanhos claros são encontrados em populações de diferentes ancestralidades (Park et al. 2012; Adhikari et al. 2016; Lock-Andersen et al. 1998), o que mostra que a associação das frequências alélicas com esses fenótipos não é tão trivial. O mesmo ocorre com o haplótipo MITF-77, também composto pelos alelos rs6761994*G, rs4855323*T e rs62250968*C, e associado a olho castanho claro.

O haplótipo MITF-52 foi associado com pele escura e, assim como os haplótipos MITF-55 e 77, possui os alelos rs6761994*G, rs4855323*T e rs62250968*C cujas frequências são predominantes no leste asiático, onde o fenótipo de pele clara é mais comum (Day et al. 2016). Adicionalmente, o alelo rs139487027*T, que diferencia o haplótipo MITF-52 dos demais, possui frequências extremamente baixas em todas as populações mundiais e não foi associado a nenhum fenótipo de pigmentação.

A pequena concordância entre as frequências alélicas mundiais das variantes apresentadas na Tabela 25 e as associações encontradas, somados ao fato de essas associações não serem significativas após a correção de Bonferroni ($\alpha = 0,0004$), é um indicativo de que

essas variantes não afetam a função e expressão do gene *MITF* de maneira a contribuir para o desenvolvimento dos diferentes fenótipos de pigmentação, e que as associações encontradas para os haplótipos MITF-52, 55 e 77 podem ser, na realidade, falsos positivos. No entanto, como dito anteriormente, interações mais complexas como epistasia, poderiam explicar as associações encontradas (Branicki et al. 2009)

Tabela 25. Frequências alélicas das variantes não associadas a fenótipos de pigmentação no gene *MITF* nas populações mundiais.

Ponto de		AFR	AMR	EUR	EAS	SAS
rs62253052	T	0,983	0,844	0,747	0,994	0,909
	A	0,017	0,156	0,253	0,006	0,091
rs9837720	G	0,799	0,785	0,944	0,925	0,864
	A	0,201	0,215	0,056	0,075	0,136
rs1427421	A	0,278	0,046	0,075	0,012	0,278
	G	0,722	0,954	0,925	0,988	0,722
rs7651891	A	0,526	0,382	0,294	0,249	0,519
	G	0,474	0,618	0,706	0,751	0,481
rs6761994	G	0,899	0,820	0,599	0,977	0,797
	A	0,101	0,180	0,401	0,023	0,203
rs4855323	G	0,950	0,764	0,765	0,553	0,809
	T	0,050	0,236	0,235	0,447	0,191
rs62250968	C	0,797	0,821	0,582	0,977	0,783
	T	0,203	0,179	0,418	0,023	0,217

8. Conclusões

8. Conclusões

Os resultados desse trabalho reafirmam a importância do gene *MITF* no desenvolvimento dos diversos fenótipos de pigmentação encontrados na população estudada. Como esperado, foram encontradas variantes que podem alterar os níveis de expressão do fator MITF e, consequentemente, a pigmentação da pele, olhos e cabelos.

Foram encontradas 21 associações fenotípicas que, até o momento, não haviam sido reportadas. Dentre essas associações, existem quatro *singletons*.

Dentre as regiões estudadas, a promotora apresenta as associações mais fortes, em especial para as variantes rs74845300 e rs80232752. De maneira geral, variantes nessa região estão associadas aos fenótipos mais pigmentados como cabelos pretos, olhos castanho escuros e pele escura.

Na região codificante, somente uma variante (*singletons* não inclusos) é exônica, porém, sinônima. Dessa forma, as variantes presentes nessa região alteram a expressão, mas não a função do fator MITF. O reduzido número de variantes exônicas indica a conservação do gene *MITF* em humanos.

A região 3'UTR apresentou o menor número de variantes e haplótipos associados a fenótipos de pigmentação. As interações miRNAs/mRNAs não puderam explicar completamente a associação encontrada para o haplótipo 3'UTR-1d1 com fenótipos escuros.

Foram encontradas, também, associações entre fenótipos de pigmentação e haplótipos das regiões do gene *MITF* e sua extensão total. Muitas dessas associações reafirmaram as encontradas para pontos de variação dentro de cada região. Em especial, o haplótipo Promo-1h1 que contém as variantes rs74845300 e rs80232752, todos associados a fenótipos escuros.

Referências

- Adhikari, K. et al., 2016. A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications*, 7, p.10815.
- Ancans, J. et al., 2001. Melanosomal pH Controls Rate of Melanogenesis, Eumelanin/Phaeomelanin Ratio and Melanosome Maturation in Melanocytes and Melanoma Cells. *Experimental Cell Research*, 268(1), p.26–35.
- Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.
- Arts, N. et al., 2015. microRNA-155, Induced by Interleukin-1 β , Represses the Expression of Microphthalmia-Associated Transcription Factor (MITF-M) in Melanoma Cells S. S. Dadras, org. *PLOS ONE*, 10(4), p.e0122517.
- Van der Auwera, G.A. et al., 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. 11.10.1-11.10.33.
- Bailey, S.D. et al., 2015. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nature Communications*, 2, p.6186.
- Balaresque, P. & King, T.E., 2016. Human Phenotypic Diversity: An Evolutionary Perspective. In *Current Topics in Developmental Biology*. Elsevier Inc., p. 349–390.
- Barrett, J.C. et al., 2005. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), p.263–265.
- Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2), p.281–97.
- Van Belle, G. et al., 2004. *Biostatistics: A Methodology for the Health Sciences*, Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Bemis, L.T. et al., 2008. MicroRNA-137 Targets Microphthalmia-Associated Transcription Factor in Melanoma Cell Lines. *Cancer Research*, 68(5), p.1362–1368.
- Bharti, K. et al., 2008. Alternative promoter use in eye development: the complex role and regulation of the transcription factor MITF. *Development (Cambridge, England)*, 135(6), p.1169–78.
- Boeva, V., 2016. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Frontiers in Genetics*, 7(FEB), p.24.
- Bonaccorso, N.S., 2005. Aplicação do Exame de Dna Na Elucidação De Crimes. *Dissertação (Mestrado) - Universidade de São Paulo*, p.1–193.
- Bonilla, C. et al., 2014. Skin pigmentation, sun exposure and vitamin D levels in children of the Avon Longitudinal Study of Parents and Children. *BMC public health*, 14(1), p.597.
- Branicki, W., Brudnik, U. & Wojas-Pelc, A., 2009. Interactions Between HERC2 , OCA2 and MC1R May Influence Human Pigmentation Phenotype. *Annals of Human Genetics*, 73(2), p.160–170.
- 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), p.2849–2856.

Castelli, E.C. et al., 2015. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Human Immunology*, 76(12), p.945–953.

Cheli, Y. et al., 2010. Fifteen-year quest for microphthalmia-associated transcription factor target genes. *Pigment Cell and Melanoma Research*, 23(1), p.27–40.

Chen, W. & Zhang, J., 2016. Potential molecular characteristics in situ in response to repetitive UVB irradiation. *Diagnostic Pathology*, 11(1), p.129.

Cher, C. (Field A.S., 2011. Bioanalyzer Applications for Next Gen Sequencing : Updates and Tips. , p.1–49.

Choudhuri, S. & Carlson, D.B., 2008. Genes and Genomes. In *Genomics: Fundamentals and Applications*. CRC Press, p. 1–1.

Cingolani, P. et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), p.80–92.

Claes, P., Liberton, D.K., et al., 2014. Modeling 3D Facial Shape from DNA. *PLoS Genetics*, 10(3).

Claes, P., Hill, H. & Shriver, M.D., 2014. Toward DNA-based facial composites: Preliminary results and validation. *Forensic Science International: Genetics*, 13, p.208–216.

Cronin, J.C. et al., 2009. Frequent mutations in the MITF pathway in melanoma. *Pigment Cell & Melanoma Research*, 22(4), p.435–444.

D'Orazio, J. et al., 2013. Melanoma — Epidemiology, Genetics and Risk Factors. In *Recent Advances in the Biology, Therapy and Management of Melanoma*. InTech, p. 3–36.

Danko, C.G. et al., 2015. Identification of active transcriptional regulatory elements from GRO-seq data. *Nature methods*, 12(5), p.433–8.

Day, A.K. et al., 2016. Acculturation, Skin Tone Preferences, and Tanning Behaviours Among Young Adult Asian Australians. *The Journal of Primary Prevention*, 37(5), p.421–432.

DePristo, M.A. et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), p.491–8.

Erlich, Y. et al., 2008. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nature Methods*, 5(8), p.679–682.

Ewing, B. et al., 1998. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*, 8(3), p.175–185.

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), p.564–567.

Excoffier, L. & Slatkin, M., 1998. Incorporating genotypes of relatives into a test of linkage disequilibrium. *American Journal of Human Genetics*, 62(1), p.171–180.

Fang, D., Tsuji, Y. & Setaluri, V., 2002. Selective down-regulation of tyrosinase family gene TYRP1 by inhibition of the activity of melanocyte transcription factor, MITF. *Nucleic acids research*, 30(14), p.3096–106.

Felix, G.E.S. et al., 2010. Ancestry informative markers and complete blood count

parameters in Brazilian blood donors. *Revista Brasileira de Hematologia e Hemoterapia*, 32(4), p.282–285.

Ferragina, P. et al., 2004. An Alphabet-Friendly FM-Index. , p.150–160.

Foley, J.W. & Sidow, A., 2013. Transcription-factor occupancy at HOT regions quantitatively predicts RNA polymerase recruitment in five human cell lines. *BMC Genomics*, 14(1), p.720.

Fuse, N. et al., 1999. Molecular Cloning of cDNA Encoding a Novel Microphthalmia-Associated Transcription Factor Isoform with a Distinct Amino-Terminus. *J Biochem*, 126(6), p.1043–1051.

Gabriel, S.B., 2002. The Structure of Haplotype Blocks in the Human Genome. *Science*, 296(5576), p.2225–2229.

George, A. et al., 2016. Biallelic Mutations in MITF Cause Coloboma, Osteopetrosis, Microphthalmia, Macrocephaly, Albinism, and Deafness. *The American Journal of Human Genetics*, 99(6), p.1388–1394.

Gotea, V. et al., 2015. The functional relevance of somatic synonymous mutations in melanoma and other cancers. *Pigment Cell & Melanoma Research*, 28(6), p.673–684.

Graffelman, J. et al., 2015. Exact Inference for Hardy-Weinberg Proportions with Missing Genotypes: Single and Multiple Imputation. *G3 (Bethesda, Md.)*, 5(11), p.2365–73.

Griffiths-Jones, S. et al., 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research*, 34(Database issue), p.D140–4.

Grill, C. et al., 2013. MITF mutations associated with pigment deficiency syndromes and melanoma have different effects on protein function. *Human Molecular Genetics*, 22(21), p.4357–4367.

Guerreiro, J. et al., 1993. Genetic Structure and Demography of the Human Population of Obidos, in the Brazilian Amazon. *Braz J Genet*, 16, p.1075–1084.

Guo, J. et al., 2014. MicroRNA-218 inhibits melanogenesis by directly suppressing microphthalmia-associated transcription factor expression. *RNA Biology*, 11(6), p.732–741.

Guo, L., Li, X. & Tang, Q.-Q., 2015. Transcriptional Regulation of Adipocyte Differentiation: A Central Role for CCAAT/Enhancer-binding Protein (C/EBP) β . *Journal of Biological Chemistry*, 290(2), p.755–761.

Haflidadóttir, B.S. et al., 2010. miR-148 Regulates Mitf in Melanoma Cells D.-Y. Jin, org. *PLoS ONE*, 5(7), p.e11574.

Hartl, D.L. & Clark, A.G., 2007. *Principles of population genetics* A. D. Sinauer, org., Sunderland, Massachussetts: Sinauer Associates, Inc. Publishers.

Hartman, M.L. & Czyz, M., 2015. MITF in melanoma: mechanisms behind its expression and activity. *Cellular and Molecular Life Sciences*, 72(7), p.1249–1260.

Hershey, C.L. & Fisher, D.E., 2005. Genomic analysis of the Microphthalmia locus and identification of the MITF-J/Mitf-J isoform. *Gene*, 347(1), p.73–82.

Hirobe, T., 2011. How are proliferation and differentiation of melanocytes regulated? *Pigment Cell and Melanoma Research*, 24(3), p.462–478.

Huntzinger, E. & Izaurralde, E., 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature reviews. Genetics*, 12(2), p.99–110.

- Jablonski, N.G. & Chaplin, G., 2010. Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences*, 107(Supplement_2), p.8962–8968.
- John, B. et al., 2004. Human MicroRNA targets. *PLoS biology*, 2(11), p.e363.
- Joyce, C.E. et al., 2011. Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. *Human Molecular Genetics*, 20(20), p.4025–4040.
- Kaikkonen, M.U. et al., 2013. Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. *Molecular Cell*, 51(3), p.310–325.
- Kayser, M., 2015. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics*, 18, p.33–48.
- Koludrovic, D. & Davidson, I., 2013. MITF, the Janus transcription factor of melanoma. *Future Oncology*, 9(2), p.235–244.
- Lassmann, T., Hayashizaki, Y. & Daub, C.O., 2011. SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics*, 27(1), p.130–131.
- Le, C.T., 2003. *Introductory Biostatistics*, Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Lee, W.-P. et al., 2016. Akt phosphorylates myc-associated zinc finger protein (MAZ), releases P-MAZ from the p53 promoter, and activates p53 transcription. *Cancer Letters*, 375(1), p.9–19.
- Di Leva, G. & Croce, C.M., 2013. MiRNA profiling of cancer. *Current Opinion in Genetics and Development*, 23(1), p.3–11.
- Levy, S.E. & Myers, R.M., 2016. Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, 17(1), p.95–115.
- Lewis, B.P. et al., 2003. 33-Prediction of Mammalian MicroRNA Targets. *Cell*, 115(7), p.787–798.
- Lewontin, R.C., 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*, 49(1), p.49–67.
- Lewontin, R.C. & Kojima, K., 1960. The Evolutionary Dynamics of Complex Polymorphisms. *Evolution*, 14(4), p.458.
- Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), p.2078–2079.
- Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), p.1754–1760.
- Li, M.J. & Wang, J., 2015. Current trend of annotating single nucleotide variation in humans – A case study on SNVrap. *Methods*, 79–80, p.32–40.
- Lock-Andersen, J., Wulf, H.C. & Knudstorp, N.D., 1998. Interdependence of eye and hair colour, skin type and skin pigmentation in a Caucasian population. *Acta dermato-venereologica*, 78(3), p.214–9.
- Lu, S.Y. et al., 2010. Subcellular localization of Mitf in monocytic cells. *Histochemistry and Cell Biology*, 133(6), p.651–658.
- Maddodi, N., Jayanthi, A. & Setaluri, V., 2012. Shining light on skin pigmentation: The

darker and the brighter side of effects of UV radiation. *Photochemistry and Photobiology*, 88(5), p.1075–1082.

Mak, S.-S. et al., 2006. Indispensable role of Bcl2 in the development of the melanocyte stem cell. *Developmental Biology*, 291(1), p.144–153.

Marin, R.M. & Vanicek, J., 2011. Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Research*, 39(1), p.19–29.

Maroñas, O. et al., 2014. Development of a forensic skin colour predictive test. *Forensic Science International: Genetics*, 13, p.34–44.

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

Maruotti, J. et al., 2012. MITF-M, a “melanocyte-specific” isoform, is expressed in the adult retinal pigment epithelium. *Pigment Cell & Melanoma Research*, 25(5), p.641–644.

Matveeva, E. et al., 2016. Involvement of PARP1 in the regulation of alternative splicing. *Cell Discovery*, 2, p.15046.

McEvoy, B., Beleza, S. & Shriver, M.D., 2006. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human molecular genetics*, 15 Spec No(Review Issue 2), p.R176–81.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), p.1297–1303.

Merkenschlager, M. & Nora, E.P., 2016. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual Review of Genomics and Human Genetics*, 17(1), p.17–43.

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), p.1215.

Mione, M. & Bosserhoff, A., 2015. MicroRNAs in melanocyte and melanoma biology. *Pigment Cell and Melanoma Research*, 28(3), p.340–354.

Musgrave-Brown, E. et al., 2007. Forensic validation of the SNPforID 52-plex assay. *Forensic Science International: Genetics*, 1(2), p.186–190.

Nan, H. et al., 2009. Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. *International Journal of Cancer*, 125(4), p.909–917.

Nishimura, E.K., 2005. Mechanisms of Hair Graying: Incomplete Melanocyte Stem Cell Maintenance in the Niche. *Science*, 307(5710), p.720–724.

Oboki, K. et al., 2002. Isoforms of mi transcription factor preferentially expressed in cultured mast cells of mice. *Biochemical and biophysical research communications*, 290, p.1250–1254.

Palmero, E.I. et al., 2011. Mechanisms and role of microRNA deregulation in cancer onset and progression. *Genetics and Molecular Biology*, 34(3), p.363–370.

Park, S.L. et al., 2012. Risk Factors for Malignant Melanoma in White and Non-White/Non-African American Populations: The Multiethnic Cohort. *Cancer Prevention Research*, 5(3), p.423–434.

Paschou, P. et al., 2010. Ancestry informative markers for fine-scale individual assignment to worldwide populations. *Journal of Medical Genetics*, 47(12), p.835–847.

Peled, S. et al., 2016. De-novo protein function prediction using DNA binding and RNA binding proteins as a test case. *Nature communications*, 7, p.13424.

Pena, S.D.J. et al., 2011. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. H. Harpending, org. *PloS one*, 6(2), p.e17063.

Pezzella, A. et al., 1996. Oxidative polymerisation of 5,6-dihydroxyindole-2-carboxylic acid to melanin: A new insight. *Tetrahedron*, 52(23), p.7913–7920.

Pezzella, A., Vogna, D. & Protà, G., 2003. Synthesis of optically active tetrameric melanin intermediates by oxidation of the melanogenic precursor 5,6-dihydroxyindole-2-carboxylic acid under biomimetic conditions. *Tetrahedron Asymmetry*, 14(9), p.1133–1140.

Pingault, V. et al., 2010. Review and update of mutations causing Waardenburg syndrome. *Human Mutation*, 31(4), p.391–406.

Porto, I.O.P. et al., 2015. MicroRNAs targeting the immunomodulatory HLA-G gene: A new survey searching for microRNAs with potential to regulate HLA-G. *Molecular Immunology*, 65(2), p.230–241.

Primot, A. et al., 2010. ERK-regulated differential expression of the Mitf 6a/b splicing isoforms in melanoma. *Pigment Cell and Melanoma Research*, 23(1), p.93–102.

Raymond, M. & Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity*, 86, p.248–249.

Rehmsmeier, M. et al., 2004. Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10), p.1507–17.

Riesz, J., 2007. The spectroscopic properties of melanin. *Tese (Doutorado) - University of Queensland*, p.1–239.

Robinson, J.T. et al., 2011. Integrative genomics viewer. *Nature Biotechnology*, 29(1), p.24–26.

Rodrigues de Moura, R. et al., 2015. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *American Journal of Human Biology*, 27(5), p.674–680.

Rossberg, W. et al., 2016. Human Pigmentation, Cutaneous Vitamin D Synthesis and Evolution: Variants of Genes (SNPs) Involved in Skin Pigmentation Are Associated with 25(OH)D Serum Concentration. *Anticancer research*, 36(3), p.1429–1437.

Rousset, F., 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8(1), p.103–106.

Ruiz-Linares, A. et al., 2014. Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. *PLoS Genetics*, 10(9).

Rumel, D., 1986. “Odds Ratio”: Algumas Considerações. *Rev Saúde Pública*, 20(3), p.253–258.

Secretaria Nacional de Segurança Pública, 2014. *Investigação Criminal de Homicídios*, Brasília.

Segura, M.F. et al., 2009. Aberrant miR-182 expression promotes melanoma metastasis by repressing FOXO3 and microphthalmia-associated transcription factor. *Proceedings of the National Academy of Sciences*, 106(6), p.1814–1819.

Shao, P. et al., 2016. Histone demethylase PHF8 promotes epithelial to mesenchymal transition and breast tumorigenesis. *Nucleic acids research*, 44(11), p.3256–3270.

Shi, Q. et al., 2016. Oxidative stress-induced overexpression of miR-25: the mechanism underlying the degeneration of melanocytes in vitiligo. *Cell Death and Differentiation*, 23(3), p.496–508.

Smith, S.D. et al., 2000. Tietz syndrome (hypopigmentation/deafness) caused by mutation of MITF. *Journal of medical genetics*, 37(6), p.446–8.

Stephens, M. & Donnelly, P., 2003. A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *The American Journal of Human Genetics*, 73(5), p.1162–1169.

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), p.978–989.

Sturm, R.A., 2009. Molecular genetics of human pigmentation diversity. *Human Molecular Genetics*, 18(R1), p.R9–R17.

Sturm, R.A. & Frudakis, T.N., 2004. Eye colour: Portals into pigmentation genes and ancestry. *Trends in Genetics*, 20(8), p.327–332.

Tachibana, M., 2000. MITF: a stream flowing for pigment cells. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society*, 13, p.230–240.

Takeda, K. et al., 2000. Ser298 of MITF, a mutation site in Waardenburg syndrome type 2, is a phosphorylation site with functional significance. *Human molecular genetics*, 9(1), p.125–32.

Takemoto, C.M., Yoon, Y.-J. & Fisher, D.E., 2002. The Identification and Functional Characterization of a Novel Mast Cell Isoform of the Microphthalmia-associated Transcription Factor. *Journal of Biological Chemistry*, 277(33), p.30244–30252.

Teng, M. et al., 2012. Regsnps: A strategy for prioritizing regulatory single nucleotide substitutions. *Bioinformatics*, 28(14), p.1879–1886.

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), p.178–192.

Timofeeva, O.A. et al., 2012. Mechanisms of Unphosphorylated STAT3 Transcription Factor Binding to DNA. *Journal of Biological Chemistry*, 287(17), p.14192–14200.

Tully, G., 2007. Genotype versus phenotype: Human pigmentation. *Forensic Science International: Genetics*, 1(2), p.105–110.

Udono, T. et al., 2000. Structural organization of the human microphthalmia-associated transcription factor gene containing four alternative promoters. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1491(1–3), p.205–219.

Vachtenheim, J. & Borovanský, J., 2010. “Transcription physiology” of pigment formation in melanocytes: Central role of MITF. *Experimental Dermatology*, 19(7), p.617–627.

Velho, J.A., Geiser, G.C. & Espindula, A., 2011. Ciências Forenses - Uma introdução às principais áreas da Criminalística Moderna. In *Ciências Forenses - Uma introdução às principais áreas da Criminalística Moderna*. Millenium.

Walsh, S. et al., 2011. Developmental validation of the IrisPlex system: Determination

of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics*, 5(5), p.464–471.

Walsh, S. et al., 2013. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, 7(1), p.98–115.

Wan, P., Hu, Y. & He, L., 2011. Regulation of melanocyte pivotal transcription factor MITF by some other transcription factors. *Molecular and Cellular Biochemistry*, 354(1–2), p.241–246.

Wang, Y. et al., 2012. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nature Structural & Molecular Biology*, 19(10), p.1044–1052.

Waples, R.S., 2015. Testing for Hardy-Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity*, 106(1), p.1–19.

Ward, L.D. & Kellis, M., 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research*, 40(D1), p.D930–D934.

Ward, L.D. & Kellis, M., 2016. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Research*, 44(D1), p.D877–D881.

Winchester, C.L. et al., 1999. Characterization of the expression of DMPK and SIX5 in the human eye and implications for pathogenesis in myotonic dystrophy. *Human molecular genetics*, 8(3), p.481–492.

Yang, X. et al., 2013. HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*, 14(1), p.33.

Yang, Y., Xie, B. & Yan, J., 2014. Application of Next-generation Sequencing Technology in Forensic Science. *Genomics, Proteomics & Bioinformatics*, 12(5), p.190–197.

Yokoyama, S. et al., 2011. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature*, 480(7375), p.99–103.

Zhang, H. et al., 2012. Functional analysis of MITF gene mutations associated with Waardenburg syndrome type 2. *FEBS Letters*, 586(23), p.4126–4131.

Apêndice A – Variantes do gene *MITF*

Tabela 26. Variantes do gene *MITF* encontradas em uma população de 340 indivíduos da população brasileira

Região	ID	Posição	Posição Relativa	(continua)					
				Alelos			Frequências		
				REF	ALT 1	ALT 2	REF	ALT 1	ALT 2
Promotora	rs115736993	chr3:69785786	-2800	A	G	-	0,9985	0,0015	-
Promotora	rs200471673	chr3:69785895	-2691	AT	A	TT	0,9809	0,0147	0,0044
Promotora	rs186582147	chr3:69785919	-2667	A	G	-	0,9985	0,0015	-
Promotora	rs62253052	chr3:69786037	-2549	T	A	-	0,7838	0,2162	-
Promotora	rs9824825	chr3:69786104	-2482	T	C	-	0,9926	0,0074	-
Promotora	rs9837720	chr3:69786140	-2446	G	A	-	0,9132	0,0868	-
Promotora	rs181456408	chr3:69786272	-2314	G	A	-	0,9941	0,0059	-
Promotora	rs57953945	chr3:69786647	-1939	CT	C	-	0,9956	0,0044	-
Promotora	-	chr3:69786651	-1935	G	A	-	0,9985	0,0015	-
Promotora	rs139094245	chr3:69786693	-1893	CATGTTTCTCAATGGGG	C	-	0,9662	0,0338	-
Promotora	rs74845300	chr3:69786714	-1872	A	G	-	0,9912	0,0088	-
Promotora	rs147055911	chr3:69786943	-1643	T	A	-	0,9971	0,0029	-
Promotora	rs182103206	chr3:69787195	-1391	G	A	-	0,9971	0,0029	-
Promotora	-	chr3:69787219	-1367	G	A	-	0,9985	0,0015	-
Promotora	rs150351819	chr3:69787306	-1280	C	A	-	0,9971	0,0029	-
Promotora	rs80232752	chr3:69787573	-1013	A	T	-	0,9912	0,0088	-
Promotora	-	chr3:69787617	-969	C	T	-	0,9985	0,0015	-
Promotora	rs74342736	chr3:69787629	-957	A	T	-	0,9985	0,0015	-
Promotora	-	chr3:69787710	-876	G	A	-	0,9985	0,0015	-
Promotora	rs1427421	chr3:69787714	-872	A	G	-	0,1132	0,8868	-
Promotora	-	chr3:69787723	-863	T	C	-	0,9985	0,0015	-
Promotora	rs117760964	chr3:69787756	-830	C	A	-	0,9985	0,0015	-
Promotora	rs12489786	chr3:69788068	-518	T	C	-	0,9779	0,0221	-
Promotora	-	chr3:69788091	-495	G	C	-	0,9985	0,0015	-
Promotora	rs77527948	chr3:69788197	-389	G	A	-	0,9985	0,0015	-
Promotora	rs7651891	chr3:69788321	-265	A	G	-	0,3353	0,6647	-
Promotora	rs531268453	chr3:69788550	-36	C	T	-	0,9985	0,0015	-
Promotora	rs114935806	chr3:69788560	-26	C	T	-	0,9971	0,0029	-
Promotora	-	chr3:69788567	-19	C	T	-	0,9971	0,0029	-
Promotora	rs561386441	chr3:69788574	-12	C	G	-	0,9941	0,0059	-
5'UTR	rs530113723	chr3:69788643	57	A	C	-	0,9971	0,0029	-
5'UTR	-	chr3:69788684	98	A	G	-	0,9985	0,0015	-
5'UTR	-	chr3:69788724	138	A	G	-	0,9985	0,0015	-
Íntron	-	chr3:69801082	12496	T	C	-	0,9985	0,0015	-
Íntron	rs6761994	chr3:69801131	12545	G	A	-	0,6838	0,3162	-
Íntron	rs113085085	chr3:69801287	12701	A	C	-	0,9985	0,0015	-
Íntron	rs13322505	chr3:69801975	13389	T	C	-	0,9897	0,0103	-
Íntron	rs185794091	chr3:69802006	13420	C	T	-	0,9985	0,0015	-
Íntron	rs72950030	chr3:69807749	19163	C	T	-	0,9824	0,0176	-
Íntron	rs76992570	chr3:69807750	19164	G	A	-	0,9971	0,0029	-
Íntron	-	chr3:69807846	19260	A	G	-	0,9985	0,0015	-

Tabela 26. Variantes do gene MITF encontradas em uma população de 340 indivíduos da população brasileira

(continuação)

Região	ID	Posição	Posição Relativa	Alelos			Frequências		
				REF	ALT 1	ALT 2	REF	ALT 1	ALT 2
Íntron	rs4855323	chr3:69807905	19319	G	T	-	0,7824	0,2176	-
Íntron	-	chr3:69807915	19329	A	G	-	0,9985	0,0015	-
Íntron	-	chr3:69807965	19379	C	T	-	0,9985	0,0015	-
Íntron	rs62250968	chr3:69811900	23314	C	T	-	0,6485	0,3515	-
Íntron	-	chr3:69812747	24161	G	A	-	0,9985	0,0015	-
Íntron	-	chr3:69812798	24212	C	T	-	0,9985	0,0015	-
Íntron	-	chr3:69812835	24249	G	A	-	0,9985	0,0015	-
Íntron	rs72950037	chr3:69812844	24258	G	T	-	0,9897	0,0103	-
Íntron	rs80163893	chr3:69812860	24274	C	T	-	0,9956	0,0044	-
Íntron	-	chr3:69812868	24282	T	A	-	0,9985	0,0015	-
Íntron	rs72950039	chr3:69812903	24317	C	T	-	0,9868	0,0132	-
Íntron	-	chr3:69827952	39366	A	AT	-	0,9985	0,0015	-
Íntron	rs2217456	chr3:69887578	98992	A	G	-	0,4745	0,5255	-
Íntron	rs1367370	chr3:69887656	99070	A	G	-	0,8176	0,1824	-
Íntron	rs1367371	chr3:69887699	99113	C	A	-	0,4691	0,5309	-
Íntron	rs545897666	chr3:69887767	99181	T	C	-	0,9956	0,0044	-
Íntron	rs78124318	chr3:69887807	99221	G	A	-	0,9588	0,0412	-
Íntron	-	chr3:69887951	99365	T	C	-	0,9985	0,0015	-
Íntron	-	chr3:69887983	99397	G	C	-	0,9985	0,0015	-
Íntron	-	chr3:69887990	99404	G	A	-	0,9985	0,0015	-
Íntron	-	chr3:69887991	99405	C	T	-	0,9985	0,0015	-
Íntron	rs141467928	chr3:69928273	139687	G	A	-	0,9985	0,0015	-
Éxon	rs9849776	chr3:69928510	139924	G	A	-	0,9912	0,0088	-
Íntron	-	chr3:69928557	139971	G	C	-	0,9985	0,0015	-
Íntron	rs112598669	chr3:69985718	197132	G	A	-	0,9926	0,0074	-
Íntron	-	chr3:69985742	197156	C	T	-	0,9985	0,0015	-
Íntron	rs77588960	chr3:69985838	197252	G	A	-	0,9750	0,0250	-
Íntron	rs199515469	chr3:69985860	197274	C	T	-	0,9985	0,0015	-
Íntron	rs372701105	chr3:69985924	197338	T	C	-	0,9985	0,0015	-
Íntron	rs189927456	chr3:69985973	197387	T	C	-	0,9956	0,0044	-
Íntron	rs73117325	chr3:69986144	197558	A	G	-	0,9838	0,0162	-
Íntron	-	chr3:69986163	197577	T	A	-	0,9985	0,0015	-
Íntron	-	chr3:69986354	197768	TA	T	-	0,9971	0,0029	-
Éxon	rs140663277	chr3:69986984	198398	C	T	-	0,9971	0,0029	-
Íntron	rs113394882	chr3:69987326	198740	C	T	-	0,9838	0,0162	-
Íntron	-	chr3:69987405	198819	AC	A	-	0,9985	0,0015	-
Íntron	rs78545229	chr3:69987480	198894	C	T	-	0,9750	0,0250	-
Íntron	rs145438735	chr3:69987691	199105	G	A	-	0,9956	0,0044	-
Íntron	rs113310138	chr3:69987707	199121	C	T	-	0,9544	0,0456	-
Íntron	rs188710184	chr3:69987763	199177	T	C	-	0,9971	0,0029	-
Íntron	-	chr3:69987778	199192	CT	C	-	0,9985	0,0015	-
Íntron	-	chr3:69987914	199328	A	G	-	0,9971	0,0029	-
Íntron	rs374067688	chr3:69988337	199751	T	C	-	0,9971	0,0029	-
Íntron	rs201698367	chr3:69988347	199761	G	C	-	0,9985	0,0015	-

Tabela 26. Variantes do gene MITF encontradas em uma população de 340 indivíduos da população brasileira

(continuação)

Região	ID	Posição	Posição Relativa	Alelos			Frequências		
				REF	ALT 1	ALT 2	REF	ALT 1	ALT 2
Íntron	rs28583802	chr3:69990278	201692	A	G	-	0,9912	0,0088	-
Íntron	rs79668299	chr3:69990714	202128	C	T	-	0,9926	0,0074	-
Íntron	-	chr3:69998393	209807	G	A	-	0,9971	0,0029	-
Íntron	-	chr3:69998442	209856	T	C	-	0,9956	0,0044	-
Íntron	rs186002522	chr3:69998476	209890	C	T	-	0,9971	0,0029	-
Íntron	-	chr3:70000955	212369	G	A	-	0,9985	0,0015	-
Íntron	-	chr3:70000957	212371	G	A	-	0,9956	0,0044	-
Íntron	rs201353723	chr3:70005695	217109	C	T	-	0,9985	0,0015	-
Íntron	rs55754687	chr3:70005712	217126	T	C	-	0,9779	0,0221	-
Éxon	rs576634776	chr3:70008502	219916	C	T	-	0,9985	0,0015	-
Íntron	rs114827487	chr3:70008665	220079	C	T	-	0,9956	0,0044	-
Íntron	-	chr3:70008701	220115	A	C	-	0,9985	0,0015	-
Éxon	rs149617956	chr3:70014091	225505	G	A	-	0,9985	0,0015	-
Éxon	rs78962087	chr3:70014199	225613	C	A	-	0,9985	0,0015	-
Éxon	-	chr3:70014320	225734	C	T	-	0,9985	0,0015	-
Éxon	rs36118030	chr3:70014384	225798	G	A	-	0,9956	0,0044	-
3'UTR	rs201375960	chr3:70014411	225825	T	C	-	0,9985	0,0015	-
3'UTR	rs9863910	chr3:70014447	225861	T	C	-	0,8750	0,1250	-
3'UTR	rs552044239	chr3:70014778	226192	T	C	-	0,9956	0,0044	-
3'UTR	-	chr3:70014804	226218	GA	G	-	0,9956	0,0044	-
3'UTR	-	chr3:70014885	226299	G	A	-	0,9985	0,0015	-
3'UTR	-	chr3:70014955	226369	T	C	-	0,9926	0,0074	-
3'UTR	rs146187921	chr3:70014965	226379	G	GA	GAA	0,7926	0,0074	0,2000
3'UTR	-	chr3:70014999	226413	C	G	-	0,9971	0,0029	-
3'UTR	rs201395752	chr3:70015214	226628	CT	C	-	0,9985	0,0015	-
3'UTR	rs114659118	chr3:70015223	226637	T	A	-	0,9956	0,0044	-
3'UTR	-	chr3:70015253	226667	G	A	-	0,9985	0,0015	-
3'UTR	-	chr3:70015373	226787	G	T	-	0,9985	0,0015	-
3'UTR	rs150314710	chr3:70015549	226963	T	C	-	0,9824	0,0176	-
3'UTR	-	chr3:70015555	226969	G	T	-	0,9985	0,0015	-
3'UTR	rs2131025	chr3:70015647	227061	T	C	-	0,9206	0,0794	-
3'UTR	rs80212793	chr3:70015890	227304	A	G	-	0,9985	0,0015	-
3'UTR	rs144336708	chr3:70016274	227688	G	A	-	0,9926	0,0074	-
3'UTR	rs192514030	chr3:70016298	227712	A	C	-	0,9985	0,0015	-
3'UTR	-	chr3:70016449	227863	C	T	-	0,9985	0,0015	-
3'UTR	rs576	chr3:70016467	227881	C	T	-	0,7353	0,2647	-
3'UTR	-	chr3:70016468	227882	G	A	-	0,9985	0,0015	-
3'UTR	rs78240629	chr3:70016569	227983	A	G	-	0,9956	0,0044	-
3'UTR	-	chr3:70016781	228195	C	A	-	0,9985	0,0015	-
3'UTR	-	chr3:70016822	228236	G	A	-	0,9985	0,0015	-
3'UTR	rs77962238	chr3:70016857	228271	G	A	-	0,9824	0,0176	-
3'UTR	rs704246	chr3:70016904	228318	C	T	-	0,7603	0,2397	-
3'UTR	rs139487027	chr3:70016979	228393	C	T	-	0,9897	0,0103	-
3'UTR	-	chr3:70017124	228538	A	G	-	0,9985	0,0015	-

Tabela 26. Variantes do gene MTF encontradas em uma população de 340 indivíduos da população brasileira

(conclusão)									
Região	ID	Posição	Posição Relativa	Alelos			Frequências		
				REF	ALT 1	ALT 2	REF	ALT 1	ALT 2
3'UTR	-	chr3:70017154	228568	G	A	-	0,9971	0,0029	-
3'UTR	rs180716228	chr3:70017264	228678	C	G	-	0,9985	0,0015	-
3'UTR	rs151043552	chr3:70017323	228737	C	T	-	0,9956	0,0044	-
3'UTR	rs139770177	chr3:70017432	228846	A	G	-	0,9985	0,0015	-

Apêndice B – Tabela 10 completa: Haplótipos encontrados para a extensão total do gene *MITF*

Tabela 10 (completa). Haplótipos encontrados para a extensão total do gene *MITF* considerando 68 pontos de variação em 244 amostras. Haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

(continua)

Haplótipo	Promotora	5'UTR + CD	3'UTR	Frequência
MITF-1	Promo-1	5'CD-2	3UTR-1	0,0738
MITF-2	Promo-1a	5'CD-2	3UTR-1	0,0512
MITF-3	Promo-1	5'CD-1a	3UTR-1	0,0471
MITF-4	Promo-1b	5'CD-1	3UTR-1	0,0451
MITF-5	Promo-1	5'CD-1	3UTR-1	0,0410
MITF-6	Promo-1b1	5'CD-1	3UTR-1a	0,0389
MITF-7	Promo-1a	5'CD-1	3UTR-1	0,0328
MITF-8	Promo-1b2	5'CD-1	3UTR-1	0,0328
MITF-9	Promo-1	5'CD-2	3UTR-1a1	0,0307
MITF-10	Promo-1	5'CD-3	3UTR-1	0,0307
MITF-11	Promo-1	5'CD-1	3UTR-2	0,0225
MITF-12	Promo-1	5'CD-3	3UTR-2	0,0184
MITF-13	Promo-1b	5'CD-2	3UTR-1	0,0143
MITF-14	Promo-1c	5'CD-3a1	3UTR-1	0,0143
MITF-15	Promo-1a	5'CD-2a	3UTR-1a1	0,0123
MITF-16	Promo-1b1	5'CD-2b	3UTR-1a1	0,0123
MITF-17	Promo-1a	5'CD-3	3UTR-1	0,0102
MITF-18	Promo-1b	5'CD-3	3UTR-2	0,0102
MITF-19	Promo-1b1	5'CD-1c	3UTR-1	0,0102
MITF-20	Promo-1b3	5'CD-1	3UTR-2	0,0102
MITF-21	Promo-1d	5'CD-1	3UTR-1	0,0102
MITF-22	Promo-1e	5'CD-1	3UTR-2	0,0102
MITF-23	Promo-1g	5'CD-2	3UTR-1	0,0102
MITF-24	Promo-1h1	5'CD-1	3UTR-2	0,0102
MITF-25	Promo-1	5'CD-2	3UTR-1a1a	0,0082
MITF-26	Promo-1	5'CD-2c	3UTR-2	0,0082
MITF-27	Promo-1a	5'CD-1	3UTR-2	0,0082
MITF-28	Promo-1a	5'CD-2	3UTR-1a1	0,0082
MITF-29	Promo-1b	5'CD-1	3UTR-2	0,0082
MITF-30	Promo-1b1	5'CD-2	3UTR-1	0,0082
MITF-31	Promo-1b1a	5'CD-2	3UTR-1	0,0082
MITF-32	Promo-1b2	5'CD-1	3UTR-2	0,0082
MITF-33	Promo-1c	5'CD-3a	3UTR-1	0,0082
MITF-34	Promo-1g	5'CD-1	3UTR-1	0,0082
MITF-35	Promo-1g	5'CD-2b	3UTR-2	0,0082
MITF-36	Promo-1	5'CD-1	3UTR-1a	0,0061
MITF-37	Promo-1	5'CD-1e1a	3UTR-2	0,0061
MITF-38	Promo-1	5'CD-2	3UTR-2	0,0061
MITF-39	Promo-1	5'CD-2k1a1	3UTR-1	0,0061
MITF-40	Promo-1a	5'CD-2a	3UTR-1c	0,0061
MITF-41	Promo-1b	5'CD-1e	3UTR-1	0,0061

Tabela 10 (completa). Haplótipos encontrados para a extensão total do gene *MITF* considerando 68 pontos de variação em 244 amostras. Haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

(continuação)

Haplótipo	Promotora	5'UTR + CD	3'UTR	Frequência
MITF-42	Promo-1b1	5'CD-1	3UTR-1	0,0061
MITF-43	Promo-1b1	5'CD-1	3UTR-2	0,0061
MITF-44	Promo-1c	5'CD-1	3UTR-1	0,0061
MITF-45	Promo-1d1	5'CD-2k1a1	3UTR-1	0,0061
MITF-46	Promo-1e	5'CD-1	3UTR-1	0,0061
MITF-47	Promo-1g	5'CD-1	3UTR-1b	0,0061
MITF-48	Promo-1	5'CD-1d	3UTR-2	0,0041
MITF-49	Promo-1	5'CD-2	3UTR-1a	0,0041
MITF-50	Promo-1	5'CD-2b	3UTR-1	0,0041
MITF-51	Promo-1	5'CD-2c1	3UTR-1	0,0041
MITF-52	Promo-1	5'CD-3	3UTR-1b	0,0041
MITF-53	Promo-1	5'CD-3a2a	3UTR-1d2a	0,0041
MITF-54	Promo-1	5'CD-3c	3UTR-1b	0,0041
MITF-55	Promo-1a	5'CD-1a1	3UTR-1e	0,0041
MITF-56	Promo-1a	5'CD-1b	3UTR-1	0,0041
MITF-57	Promo-1a	5'CD-2	3UTR-2	0,0041
MITF-58	Promo-1a	5'CD-2f	3UTR-1	0,0041
MITF-59	Promo-1a	5'CD-3b	3UTR-1	0,0041
MITF-60	Promo-1a	5'CD-3d	3UTR-1	0,0041
MITF-61	Promo-1b	5'CD-1	3UTR-2a	0,0041
MITF-62	Promo-1b	5'CD-1b	3UTR-2	0,0041
MITF-63	Promo-1b	5'CD-1g1	3UTR-1a1	0,0041
MITF-64	Promo-1b	5'CD-1h	3UTR-1	0,0041
MITF-65	Promo-1b1	5'CD-1	3UTR-1f	0,0041
MITF-66	Promo-1b1	5'CD-2 ^a	3UTR-1a1	0,0041
MITF-67	Promo-1b1	5'CD-3b	3UTR-1	0,0041
MITF-68	Promo-1b2	5'CD-1	3UTR-2c	0,0041
MITF-69	Promo-1b2	5'CD-2c21	3UTR-1	0,0041
MITF-70	Promo-1b3	5'CD-1	3UTR-1a	0,0041
MITF-71	Promo-1b3	5'CD-2h	3UTR-1	0,0041
MITF-72	Promo-1b3	5'CD-3	3UTR-1	0,0041
MITF-73	Promo-1c	5'CD-3	3UTR-2	0,0041
MITF-74	Promo-1f	5'CD-1d	3UTR-1	0,0041
MITF-75	Promo-1g	5'CD-2c1	3UTR-1d2a	0,0041
MITF-76	Promo-1g	5'CD-3	3UTR-1	0,0041
MITF-77	Promo-1g	5'CD-3	3UTR-2	0,0041
MITF-78	Promo-1	5'CD-1	3UTR-1d	0,0020
MITF-79	Promo-1	5'CD-1f	3UTR-1	0,0020
MITF-80	Promo-1	5'CD-1g1	3UTR-1	0,0020
MITF-81	Promo-1	5'CD-2c	3UTR-1	0,0020
MITF-82	Promo-1	5'CD-2d	3UTR-1	0,0020
MITF-83	Promo-1	5'CD-2k1	3UTR-1d2a1	0,0020
MITF-84	Promo-1	5'CD-3	3UTR-1a1	0,0020
MITF-85	Promo-1	5'CD-3	3UTR-1e	0,0020

Tabela 10 (completa). Haplótipos encontrados para a extensão total do gene *MITF* considerando 68 pontos de variação em 244 amostras. Haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

(continuação)

Haplótipo	Promotora	5'UTR + CD	3'UTR	Frequência
MITF-86	Promo-1	5'CD-3a1	3UTR-1	0,0020
MITF-87	Promo-1	5'CD-3a2a	3UTR-1	0,0020
MITF-88	Promo-1	5'CD-3b	3UTR-2	0,0020
MITF-89	Promo-1	5'CD-3e	3UTR-2	0,0020
MITF-90	Promo-1a	5'CD-2b	3UTR-1a1	0,0020
MITF-91	Promo-1a	5'CD-2h	3UTR-1	0,0020
MITF-92	Promo-1a	5'CD-2k1	3UTR-1d2a1	0,0020
MITF-93	Promo-1a	5'CD-3 ^a	3UTR-1	0,0020
MITF-94	Promo-1a	5'CD-3a1	3UTR-1	0,0020
MITF-95	Promo-1a1	5'CD-3h1	3UTR-1	0,0020
MITF-96	Promo-1b	5'CD-1	3UTR-1a	0,0020
MITF-97	Promo-1b	5'CD-1i	3UTR-1	0,0020
MITF-98	Promo-1b	5'CD-2	3UTR-1a1	0,0020
MITF-99	Promo-1b	5'CD-2	3UTR-2a	0,0020
MITF-100	Promo-1b	5'CD-2h	3UTR-1	0,0020
MITF-101	Promo-1b	5'CD-3a1	3UTR-1	0,0020
MITF-102	Promo-1b	5'CD-3b	3UTR-1a1	0,0020
MITF-103	Promo-1b1	5'CD-1c	3UTR-1a1	0,0020
MITF-104	Promo-1b1	5'CD-1k1	3UTR-1d2a1	0,0020
MITF-105	Promo-1b1	5'CD-2c1	3UTR-1	0,0020
MITF-106	Promo-1b1b	5'CD-2c1a1a1a	3UTR-2	0,0020
MITF-107	Promo-1b1b	5'CD-3c1a1a	3UTR-1d	0,0020
MITF-108	Promo-1b2	5'CD-1	3UTR-2b	0,0020
MITF-109	Promo-1b2	5'CD-1e2a	3UTR-1	0,0020
MITF-110	Promo-1b2	5'CD-1g1	3UTR-1	0,0020
MITF-111	Promo-1b2	5'CD-2c1	3UTR-1	0,0020
MITF-112	Promo-1b2	5'CD-2c2	3UTR-2	0,0020
MITF-113	Promo-1b2	5'CD-2c21	3UTR-1d2a1	0,0020
MITF-114	Promo-1b2	5'CD-3a2a	3UTR-1d2a	0,0020
MITF-115	Promo-1b2	5'CD-3c1a1	3UTR-1	0,0020
MITF-116	Promo-1b2a	5'CD-1	3UTR-1a1	0,0020
MITF-117	Promo-1b2a	5'CD-1d	3UTR-1	0,0020
MITF-118	Promo-1b2a	5'CD-2	3UTR-1	0,0020
MITF-119	Promo-1b2a	5'CD-2c21	3UTR-1	0,0020
MITF-120	Promo-1b3	5'CD-1d	3UTR-1	0,0020
MITF-121	Promo-1b3	5'CD-1f	3UTR-1	0,0020
MITF-122	Promo-1b3	5'CD-2e	3UTR-1	0,0020
MITF-123	Promo-1c	5'CD-1a1	3UTR-1	0,0020
MITF-124	Promo-1c	5'CD-2d	3UTR-1	0,0020
MITF-125	Promo-1c	5'CD-3a2a	3UTR-1d2a	0,0020
MITF-126	Promo-1d	5'CD-1	3UTR-1a1	0,0020
MITF-127	Promo-1d	5'CD-1	3UTR-2	0,0020
MITF-128	Promo-1f	5'CD-1f	3UTR-1	0,0020
MITF-129	Promo-1g	5'CD-1 ^a	3UTR-2	0,0020

Tabela 10 (completa). Haplótipos encontrados para a extensão total do gene *MITF* considerando 68 pontos de variação em 244 amostras. Haplótipos associados a algum fenótipo de pigmentação estão destacados em vermelho.

(conclusão)

Haplótipo	Promotora	5'UTR + CD	3'UTR	Frequência
MITF-130	Promo-1g	5'CD-2c1a1a1a	3UTR-1	0,0020
MITF-131	Promo-1h1	5'CD-1e2a	3UTR-1d1	0,0020
MITF-132	Promo-1i1	5'CD-2l1a	3UTR-2	0,0020

Anexo A - 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 custo. Visando lhe proporcionar maior conforto e comodidade, a coleta será realizada uma única vez em data e local a combinar.
- b) Não haverá nenhuma forma de reembolso em dinheiro, uma vez que sua participação na pesquisa não proporcionará nenhum gasto.
- c) Você terá a garantia de receber a resposta a qualquer pergunta ou esclarecimento a qualquer dúvida acerca dos procedimentos, riscos, benefícios e outros relacionados com a pesquisa. Isso poderá ser feito a qualquer momento. Nós temos o compromisso de proporcionar informações atualizadas durante o estudo, ainda que esta possa influenciar a sua vontade de continuar autorizando a sua participação.

- d) Você terá a liberdade de retirar o seu consentimento a qualquer momento sem que isso lhe traga prejuízo algum ou qualquer tipo de constrangimento.
- e) Você terá garantida a segurança de que não será identificado e que será mantida a sua privacidade e o caráter confidencial das informações concedidas.
- f) A amostra biológica será coletada com total responsabilidade e custo dos pesquisadores. A coleta de seu sangue será única e feita com material descartável, sendo necessário apenas um pequeno furo com agulha em veia do seu braço, o que pode, de fato, ser um procedimento desconfortável, mas apenas pela “picada” com a agulha.
- g) O sangue doado por você não será empregado na realização de pesquisas de outra natureza sem sua autorização.
- h) Será aplicado um questionário visando uma melhor compreensão de sua origem e características físicas. Toda informação obtida será mantida em sigilo.
- i) Os resultados da pesquisa serão apresentados em reuniões científicas (simpósios e congressos) e serão publicados em revistas científicas especializadas.

Assinatura do pesquisador responsável

Eu, _____, RG n o _____,
residente na _____ n°
_____, 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

**Anexo B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
PARA ARMAZENAMENTO DE MATERIAL BIOLÓGICO**

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

Pesquisador responsável: Prof. Dr. Celso Teixeira Mendes Junior

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

Prezado (a) doador (a),

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

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

- a) Você tem a liberdade de decidir se seu material será armazenado.
- b) Você terá a liberdade de retirar o seu consentimento a qualquer momento sem que isso lhe traga prejuízo algum ou qualquer tipo de constrangimento.
- c) O armazenamento não lhe trará qualquer custo. A amostra biológica será armazenada com total responsabilidade e custo dos pesquisadores.
- d) Toda nova pesquisa a ser realizada com o material armazenado será submetida para aprovação do Comitê de Ética em Pesquisa (CEP) institucional e, quando for o caso, da Comissão Nacional de Ética em Pesquisa (CONEP).
- e) Caso deseje, você tem o direito de ser consultado sempre que houver intenção de uso de seu material biológico em pesquisas futuras, para que eu autorize ou não autorize sua nova utilização.
- f) Você terá a garantia de receber a resposta a qualquer pergunta ou esclarecimento a qualquer dúvida acerca do seu material biológico, condições de armazenamento e sua utilização em outros projetos de pesquisa. Isso poderá ser feito a qualquer momento. Nós temos o compromisso de proporcionar informações atualizadas durante o estudo, ainda que esta possa influenciar a sua vontade de continuar autorizando a sua participação.
- g) Você terá garantida a segurança de que não será identificado e que será mantida a sua privacidade e o caráter confidencial das informações concedidas. Toda informação obtida será mantida em sigilo.
- h) Você tem o direito de ser informado sobre eventual perda ou destruição de suas amostras biológicas, bem como sobre eventual encerramento do Biorrepositório.

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

Assinatura do pesquisador responsável

Eu, _____, RG no _____, residente na _____ nº _____, cidade de _____, 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

**Anexo C - 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.

<u>1. EXTRAÇÃO DE DNA GENÔMICO</u>	<u>143</u>
<u>2. QUANTIFICAÇÃO E VERIFICAÇÃO DA QUALIDADE DO DNA</u>	<u>143</u>
<u>3. PREPARO DOS REAGENTES PARA ENRIQUECIMENTO DE ALVO</u>	<u>144</u>
<u>4. ENRIQUECIMENTO DO ALVO E PREPARO DA BIBLIOTECA</u>	<u>144</u>
4.1. Digestão e hibridação do DNA	145
4.2. Captura, ligação, eluição e PCR	146
4.3. Purificação da biblioteca	147
<u>5. CONTROLE DE QUALIDADE E QUANTIFICAÇÃO DA BIBLIOTECA</u>	<u>147</u>
<u>6. SEQUENCIAMENTO</u>	<u>148</u>
<u>7. REFERÊNCIAS</u>	<u>149</u>

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 et al. 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 ressuspenso 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 biotiniladas cujos finais são complementares a esses fragmentos-alvo (durante esse

passo os fragmentos são circularizados e índices, que irão identificar cada indivíduo durante o sequenciamento, são adicionados); 3) captura dos fragmentos-alvo utilizando *beads* magnéticas recobertas por estreptavidina; 4) ligação dos fragmentos circularizados; 5) amplificação por PCR dos fragmentos-alvo capturados para montagem da biblioteca.

4.1. Digestão e hibridação do DNA

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

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

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

O DNA digerido de cada amostra (80µL somando todos os 8 tubos de digestão) é então transferido para os microtubos contendo o mix de hibridação (70µL) e seus índices correspondentes (10µL). Essas soluções são incubadas a 54°C durante 16 horas. O tempo de duração da hibridação é determinado de acordo com a densidade de sondas do ensaio. Segundo instruções do fabricante (*Haloplex, Agilent Technologies*), ensaios que atingem cobertura de entre 500kb - 5Mb exigem 16 horas de hibridação, enquanto que ensaios com cobertura de entre 1-500kb exigem apenas 3 horas de hibridação. Apesar de o presente ensaio estar voltado para a

cobertura de apenas 488,7kb, os ensaios iniciais demonstraram a necessidade de um tempo de hibridação de 16 horas.

4.2. Captura, ligação, eluição e PCR

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

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

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

PCR: Com o auxílio da placa magnética, coletamos o DNA eluído num total de 20µL por amostra e o transferimos para tubos contendo mix de PCR. Tal mix para a reação de PCR é preparado com 209,3µL de água MilliQ, 130µL de *5X Herculase II Reaction Buffer*, 5,2µL de dNTPs (100mM, 25mM para cada dNTP), 13µL de cada um dos *primers forward* e *reverse* (25µM), 6,5µL de ácido acético (2M) e 13µL da polimerase *Herculase II Fusion*, somando um total de 30µL por amostra. O volume final da reação é de 50µL por amostra. As condições de ciclagem da PCR seguem uma etapa inicial de desnaturação por 10 minutos a 98°C, seguida por 20 ciclos de 98°C por 30 segundos, 60°C por 30 segundos e 72°C por 1 minuto, e um passo final a 72°C por 10 minutos. Após o término da reação de PCR o produto é mantido a 8°C.

4.3. Purificação da biblioteca

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

5. CONTROLE DE QUALIDADE E QUANTIFICAÇÃO DA BIBLIOTECA

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

O controle de qualidade das bibliotecas é feito através do sistema de eletroforese capilar *2100 Bioanalyzer* (*Agilent Technologies*) utilizando um chip de alta-sensibilidade. A eletroforese é capaz de separar os fragmentos com base no tamanho, permitindo dessa forma avaliar a qualidade das bibliotecas geradas. É esperado para este ensaio que as bibliotecas apresentem um perfil de picos variando de 175 a 625pb, conforme previsto por um eletroferograma padrão (Figura 1). A partir da seleção manual da faixa de fragmentos gerados (175 a 625pb) é possível obter a concentração (em pg/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>

Anexo D - 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 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.

<u>1. ANÁLISE DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO</u>	<u>151</u>
<u>1.1. Análise primária</u>	<u>151</u>
<u>1.2. Análise secundária</u>	<u>152</u>
1.2.1. Avaliação da qualidade	152
1.2.2. Trimagem	152
1.2.3. Alinhamento ao genoma de referência	154
1.2.3.1. SAM	156
1.2.3.2. BAM	157
<u>1.3. Análise terciária</u>	<u>158</u>
1.3.1. Determinação de variantes	158
1.3.1.1. VCF	161
1.3.2. Remoção de genótipos de baixa qualidade	163
<u>2. REFERÊNCIAS</u>	<u>165</u>

1. ANÁLISE DE DADOS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

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

1.1. Análise primária

O processamento dos dados segue passos gerais pré-estabelecidos, de acordo com o objetivo pretendido, e tem início no *base calling* (ou a interpretação da imagem obtida a partir do sequenciador), quando são gerados os dados brutos do sequenciamento no formato *base call* (*.bcl). Estes dados são então disponibilizados através do ambiente de computação em nuvem da *Illumina, BaseSpace* (<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 et al. 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 2010), o qual utiliza um arquivo FastQ para gerar um relatório completo da qualidade das sequências (Figura 1a). A partir dos resultados apresentados é possível decidir por estratégias que melhorem a qualidade dos dados gerados.

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

1.2.2. Trimagem

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

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

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

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

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

```
cutadapt -e 0.2 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A
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 *mem* para o BWA-MEM.

Antes de realizar o alinhamento, porém, é preciso indexar o genoma de referência por meio do índice FM (Ferragina et al. 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:sampleP'
/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/espço 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 PIG0001 representa o nome da amostra e *sorted* indica que o arquivo BAM gerado se encontra ordenado:

```
PIG0001.sorted.bam
```

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

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

```
samtools index sorted.mapped.sample.bam
```

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

```
PIG0001.sorted.bam.bai
```

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

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

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

Por outro lado, o *UnifiedGenotyper* possui filtros menos 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.

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 \
  --dbSNP dbSNP.vcf \
  -stand_call_conf 30 \
  -stand_emit_conf 10 \
  -L targets.interval_list \
  -o output.raw.snps.indels.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
-stand_emit_conf	permite emitir determinações de genótipo com baixa qualidade como resultados filtrados
-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 (<i>Variant Call Format</i>) contendo os sítios que o algoritmo (<i>HaplotypeCaller</i> ou <i>UnifiedGenotyper</i>) 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).

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 4):

Tabela 4: 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	identifica quais filtros são aplicados aos dados
FORMAT e INFO	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 5.

Tabela 5: 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

Continuação Tabela 5: Descrição dos componentes do corpo de um arquivo VCF

Informações presentes no corpo do arquivo	Descrição das abreviações apresentadas
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 0.10.1) (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 uncertain input=input.vcf output=output.vcf
```

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

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

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., Veli, M., & Navarro, G. (2004). An Alphabet-Friendly FM-Index, 150–160.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10–12. <https://doi.org/10.14806/ej.17.1.200>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In *Current Protocols in Bioinformatics* (Vol. 11, p. 11.10.1-11.10.33). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471250953.bi1110s43>

Anexo E - 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	167
2. FASEAMENTO COMPUTACIONAL	168
2.1. Faseamento completo	169
2.2. Faseamento fragmentado	170
3. DESEQUILÍBRIO DE LIGAÇÃO	171
4. FREQUÊNCIAS ALÉLICAS, EQUILÍBRIO DE HARDY-WEINBERG E HETEROZIGOSE ESPERADA	172
5. ASSOCIAÇÕES GENÓTIPO-FENÓTIPO	172
6. REFERÊNCIAS	173

1. FASEAMENTO EXPERIMENTAL

É utilizada a ferramenta *ReadBackedPhasing*, também incorporada no programa GATK (Van der Auwera et al. 2013; DePristo et al. 2011; McKenna et al. 2010), 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 (Stephens et al. 2001; 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 (Stephens et al. 2001; Stephens & Donnelly 2003).

O PHASE implementa um método de estatística Bayesiana para reconstruir os haplótipos a partir dos dados de genotipagem da população (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 et al. 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 (Rousset 2008; Raymond & Rousset 1995). 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) (Van Belle et al. 2004; Le 2003). 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., Kernysky, 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 Publica*, 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>

Anexo F - 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 (Marin & Vanicek 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 et al. 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 et al. 2004), miRanda (John et al. 2004) e intaRNA (Busch et al. 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.castelli-lab.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 predizer 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 et al. 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: $\epsilon = \epsilon_{\text{RNAhybrid}} + \epsilon_{\text{miRanda}} + \epsilon_{\text{intaRNA}}$. 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.and>