

Universidade Federal do Rio Grande do Sul
Departamento de Genética e Biologia Molecular

**Adaptação e coadaptação de genes relacionados com
pigmentação de pele e vitamina D em populações nativas americanas**

Bruna Oliveira Missaggia

Dissertação submetida ao Programa de
Pós-Graduação em Genética e Biologia
Molecular da UFRGS como requisito
parcial para a obtenção do grau de
Mestre em Genética e Biologia
Molecular

Orientadora: Profa. Dra. Maria Cátira Bortolini

Coorientadora: Profa. Dra. Tábita Hünemeier

Porto Alegre

2019

Este trabalho foi realizado no Laboratório de Evolução Humana e Molecular (LEHM) do Departamento de Genética, Instituto de Biociências da Universidade Federal do Rio Grande do Sul. O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) proveu bolsa de mestrado à aluna, vinculada ao Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM – UFRGS).

Agradecimentos

Agradeço à minha orientadora, Maria Cátira Bortolini, por ter me recebido gentilmente em seu laboratório mesmo sem me conhecer. À ela também pelo respaldo e pelo empenho com que me ajudou em todas as etapas do desenvolvimento desta dissertação.

À professora Tábita Hünemeier por ter aceitado ser minha co-orientadora e pela valiosa sugestão do tema da vitamina D em nativos americanos.

Agradeço a todos os que contribuíram com essa pesquisa, especialmente à professora Gabriela Cybis pela competente consultoria estatística, sem a qual não teria sido possível concluir a dissertação. Ao Guillermo Reales pela fundamental ajuda com os *scripts* de R e pelas proveitosas conversas sobre ciência. Ao Cainã Couto-Silva pelos *scripts* bash. E à Vanessa Jacovas pela ajuda no começo da pesquisa.

Também sou grata à Martiela Freitas pelo tempo dedicado a entender as redes. Ao Pedro Vargas por ter me recebido bem no laboratório. Ao Lucas Viscardi por ter me ensinado os primeiros comandos de bash. À Gabriela Caldas pela companhia nas disciplinas que fizemos juntas. E aos demais colegas do LEHM: Bibiana Fan, Vinícius Sortica, Gabrielle Rosa, Camila Cantele e Luane Landau pela convivência agradável no laboratório.

Agradeço também aos funcionários e professores do PPGBM pela competência e ensinamentos. Em especial ao professor Aldo Araújo que foi meu orientador durante a graduação e que muito me ensinou sobre genética.

Meus agradecimentos especiais aos que fazem parte da minha vida pessoal. Os amigos que estiveram mais presentes nesses dois anos: Alex Moraes, Juliana Mesomo, Giulia Bicca e Tobias Balbinder. E aos meus familiares, meus pais por me apoiarem em tudo e meus irmãos pela convivência e amizade. Sou grata sobretudo à minha irmã Juliana pelas conversas, conselhos e por todo o incentivo sempre.

Por fim, agradeço ao CNPQ pela bolsa concedida.

Sumário:

Resumo.....	8
Abstract.....	10
Introdução.....	12
1.1 Cor de pele e radiação ultravioleta.....	12
1.2 Pressão seletiva em ambientes com alta incidência de radiação: Fotólise da vitamina B9 e câncer de pele.....	14
1.3 Pressão seletiva em ambientes com baixa incidência de radiação: Fotossíntese de vitamina D.....	16
1.4 Evolução da cor de pele humana e a adaptação das populações nativas continentais à incidência de radiação.....	19
1.5 Variação da cor da pele em Nativos Americanos: implicações adaptativas.....	24
Objetivos.....	31
2.1 Objetivo geral.....	31
2.2 Objetivos específicos.....	31
Amostra e Métodos.....	32
3. Populações e Genes.....	32
3.1 Populações amostradas.....	32
3.2 Categorias populacionais.....	33
3.2.1 Para testar a influência das práticas de subsistência e hábitos de dieta.....	34
3.2.2 Para testar a influência da altitude.....	34
3.2.3 Para discriminar a influência das práticas de subsistência/hábitos de dieta da altitude.....	34

3.2.4 Para testar a influência da latitude.....	35
3.3 Seleção de genes e obtenção de SNPs.....	35
3.3.1 Metabolismo da vitamina D.....	35
3.3.2 Controle da expressão da vitamina D.....	35
3.3.3 Via da pigmentação da pele.....	35
4 Análises dos dados.....	36
4.1 Análises de SNPs em diferentes grupos populacionais.....	36
4.2 Análises de redes e combinações alélicas: Coeficiente de Correlação Personalizado e BlocBuster.....	37
4.2.1 Filtro de Desequilíbrio de Ligação (LD).....	38
4.2.2 Formação das redes.....	38
4.2.3 Testes de permutação e validação do limiar.....	39
4.2.4 Simulação e medida de associação das arestas.....	40
4.2.5 Comparação entre redes e combinações alélicas específicas e grupos populacionais.....	40
4.3 Estimativa do nível de estruturação das categorias.....	41
Resultados.....	42
5.1 Influência das práticas de subsistência e hábitos de dieta.....	42
5.2 Influência da altitude.....	43
5.3 Distinção entre a influência da altitude e da dieta.....	43
5.4 Influência da latitude.....	44
6. Formação das redes e medida de associação entre as interações.....	45
6.1 Filtro de LD e aplicação do BlocBuster.....	45

6.2 Testes de permutação e associação entre alelos considerando as 9 redes.....	47
6.3 Interações intergênicas (alelo-alelo).....	48
6.4 Comparação entre redes e combinações alélicas específicas e grupos populacionais.....	49
6.4.1 Influência das práticas de subsistência e hábitos de dieta.....	50
6.4.2 Influência da altitude.....	50
6.4.3 Distinção entre a influência da altitude e das práticas de subsistência/hábitos de dieta.....	51
6.4.4 Influência da latitude.....	51
7. Análises de SNPs em diferentes grupos populacionais.....	51
Discussão.....	53
8.1 Considerações sobre os métodos utilizados.....	53
8.2 Variações nos genes e vias sob investigação e suas implicações.....	57
8.2.1 Práticas de subsistência e hábitos de dieta.....	57
8.2.2 Influência da altitude.....	62
8.2.3 Influência da latitude.....	64
Conclusões e reflexões finais.....	67
Referências Bibliográficas.....	72
Imagens Suplementares.....	84
Figura suplementar 1.....	84
Tabela S1. Classificação das populações estudadas.....	85
Tabela S2. Genes envolvidos no metabolismo da vitamina D.....	89
Tabela S3. Genes envolvidos no controle da expressão da vitamina D.....	90
Tabela S4. Genes relacionados com pigmentação de pele.....	92

Tabelas suplementares 5 e 6.....	96
Tabela S7.....	97
Tabela suplementar 8.....	98

Resumo:

A adaptação humana à radiação ocorre, basicamente, pela seleção em genes relacionados à vitamina D e à pigmentação de pele. O objetivo desse trabalho é avaliar se há sinal de que um conjunto pré-selecionado de genes e variantes comuns nessas duas vias estão coevoluindo em populações nativas americanas que habitam diferentes regiões e que têm distintos hábitos de subsistência e dieta. Então podemos sugerir que, se as populações autóctones americanas estiverem adaptadas a seus respectivos contextos geográficos/ecológicos/culturais, é possível que apresentem distribuição alélica diferencial nesses *loci*.

Para testar a hipótese acima, utilizamos dados de varredura genômica, pré-selecionando SNPs de 16 genes (8 relacionados a vitamina D e 8 a pigmentação) em 333 indivíduos de 33 populações nativas. Com objetivo de testar coadaptação, aplicamos um método de redes que mede interação por meio de um cálculo de desequilíbrio de ligação.

Comparamos as frequências alélicas, bem como a frequência de indivíduos com as interações alélicas, entre as categorias populacionais relacionadas à incidência de radiação (altas altitudes vs baixas altitudes e altas latitudes vs baixas latitudes) e a potencial variação da disponibilidade de vitamina D na dieta (agriculturalistas vs caçadores-coletores). Aplicando o teste exato de Fisher com correção FDR, encontramos diferenças estatisticamente significativas em diversos SNPs, bem como em combinações alélicas que sugerem interação.

Alguns achados merecem destaque. Por exemplo, entre as categorias altas latitudes e baixas latitudes, detectamos uma variante no gene *CYP2R1* (alelo A do rs2060793), que confere mais eficiência à enzima codificada por ele, mais frequente em indivíduos de altas latitudes (e baixa incidência de UV). A maior frequência desse alelo pode ter sido favorecida pela seleção natural devido a necessidade aumentada da eficiência no metabolismo da vitamina D nesse ambiente com menos disponibilidade da vitamina.

Em relação à comparação entre as categorias altas e baixas altitudes foi encontrada em maior frequência a ligação entre alelos potencialmente

funcionais dos genes *CYP24A1* e *VDR* (alelo C do rs2248359 e alelo G do rs11574143, respectivamente) nas terras altas, o que pode ter sido resultado do relaxamento da pressão de seleção para eficiência no metabolismo desse hormônio em ambientes com alta incidência de radiação UV em populações com a pele relativamente clara.

No que diz respeito às diferenças nas práticas de subsistência e hábitos de dieta, um alelo (A do rs4516035) relacionado a maior atividade do VDR pode ter sido selecionado em agriculturistas para compensar a menor disponibilidade dessa vitamina na alimentação deles. Além disso, identificamos alelos (rs3740164/C e rs750358/T) de diferentes genes (*CUBN* e *OCA2*, respectivamente) pertencentes a vias distintas (metabolismo da vitamina D e rota da pigmentação) que podem estar sendo segregados juntos em populações nativas que praticam agricultura por favorecerem adaptativamente os indivíduos nesse contexto em particular.

Palavras-chave: Vitamina D, pigmentação de pele, nativos americanos, adaptação, coadaptação, interação gênica, radiação ultravioleta.

Abstract:

The human adaptation to UV radiation occurs, basically, by selection in genes related to vitamin D and skin pigmentation. The objective of this work is to evaluate if there is a signal that a pre-selected set of genes and common variants in these two pathways are co-evolving in Native American populations that inhabit different regions and that have distinct habits of subsistence and diet. Thus, we can suggest that if Native American populations are adapted to their respective geographic/ecological/cultural contexts, they may have differential allelic distribution at these loci.

To test the hypothesis above, we used genomic scanning data, pre-selecting SNPs from 17 genes (8 related to vitamin D and 9 to pigmentation) in 333 individuals from 33 Native American populations. In order to test co-adaptation, we applied a method of networks that measures interaction using linkage disequilibrium (LD) approaches.

We compared allele frequencies as well as the frequency their interactions, considering population categories related to UV radiation (high altitudes vs low altitudes and high latitudes vs low latitudes), as well as the potential variation of vitamin D availability in the diet (agriculturalists vs hunter-gatherers). Applying the Fisher exact test with FDR correction, we found statistically significant differences in several SNPs, as well as in allelic combinations, suggesting interaction.

Some findings deserve attention. For example, among the high latitudes and low latitudes categories, we detected a variant in the *CYP2R1* gene (allele A of rs2060793), which gives more efficiency to the enzyme encoded by it, more frequent in individuals of high latitudes (and low incidence of UV). The higher frequency of this allele may have been favored by natural selection because of the increased need for vitamin D metabolism efficiency in this environment with less vitamin availability.

Regarding the comparison between the high and low altitudes categories, the interaction between potentially functional alleles of the genes *CYP24A1* and *VDR* (rs2248359 C allele and G allele of rs11574143, respectively) was found to be more frequent in high altitudes populations, which

may have resulted from the relaxation of the selection pressure for efficiency in the metabolism of this hormone in environments with high incidence of UV radiation in populations with relatively clear skin.

Concerning differences in subsistence practices and dietary habits, an allele (A of rs4516035) related to increased VDR activity may have been selected in agriculturalists to compensate for the lower availability of this vitamin in their foods. In addition, we identified alleles (rs3740164 / C and rs750358 / T) from different genes (*CUBN* and *OCA2*, respectively) belonging to distinct pathways (vitamin D metabolism and route of pigmentation) that may be segregating together in agriculturalists, favoring adaptatively individuals from that context.

Keywords: Vitamin D, skin pigmentation, Native American, adaptation, co-adaptation, gene interaction, ultraviolet radiation.

Introdução:

1.1 Cor de pele e radiação ultravioleta

A cor de pele é uma das características fenotípicas que distingue de modo mais marcante as populações humanas. Isso é reflexo da grande diversidade interpopulacional para este traço, aproximadamente 88%, em contraposição à relativamente baixa diversidade genética geral existente entre grupos humanos, estimada entre 5% e 15% (Relethford, 2002). Além disso, há assinatura de seleção, tanto positiva como purificadora, em muitos genes envolvidos com pigmentação de pele (Hancock *et al.*, 2011; Sturm e Duffy, 2012), o que denota sua importância adaptativa.

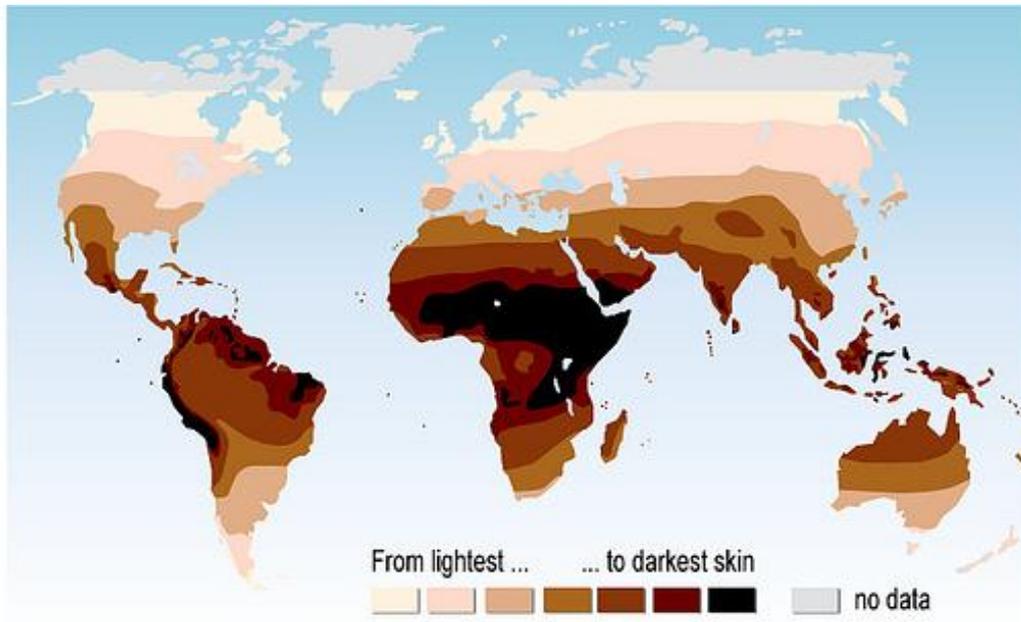
A pigmentação cutânea é determinada pela melanina, um polímero orgânico derivado do aminoácido tirosina, produzida nos melanócitos, células do estrato basal da epiderme. Existem dois tipos de melanina quimicamente distintos, a eumelanina (com coloração de marrom a preta e formato oval) e a feomelanina (com coloração de amarelo a vermelho e formato esférico) (Ito e Wakamatsu, 2003; Sturm, 2006; Scherer e Kumar, 2010; Cerqueira, *et al.*, 2014). Ambos os tipos de melanina são sintetizados e armazenados nos melanosomos, organelas tecido-específicas relacionadas aos lisossomos (Raposo *et al.*, 2009; Scherer e Kumar, 2010). Quando maduros, os melanosomos são transportados através de projeções dendríticas dos melanócitos até os queratinócitos (Sturm, 2006; Scherer e Kumar, 2010). Os diferentes tons de pele são influenciados pela quantidade, tamanho e o tipo de melanina dos melanosomos depositadas nos queratinócitos da epiderme (Barsh, 2003; Scherer e Kumar, 2010; Deng e Xu, 2018; Figura suplementar 1).

O pigmento melanina exerce um papel fotoprotetivo, bloqueando a penetração da radiação ultravioleta (RUV) na pele (Jablonski, 2004; D’Orazio, *et al.*, 2013). Atualmente acredita-se que a diversidade de pigmentação da pele em populações humanas evoluiu por uma demanda adaptativa para regular a quantidade de raios ultravioleta que penetram na pele em diferentes ambientes (Jablonski e Chaplin, 2000, 2010; McEvoy *et al.*, 2006; Lewis *et al.*, 2013; Deng e Xu, 2018, Crawford *et al.*, 2017; Martin *et al.*, 2017).

Os raios UV são a porção do espectro solar que compreende os comprimentos de onda entre 100 nm e 400 nm, entre a luz visível e os raios gama. Eles são subdivididos em três componentes com base em propriedades eletrofísicas, UVA, UVB e UVC. Os fótons UVA (400-315 nm) são os menos energéticos, como consequência de sua menor frequência têm maior poder de penetração, correspondendo entre 90% e 95% da radiação ultravioleta que penetra na atmosfera (D’Orazio, *et al.*, 2013). Os raios UVB (315-280 nm), equivalem de 5% a 10% da radiação incidente na terra, devido ao menor comprimento de onda sua penetração varia mais intensamente com a altitude e o ângulo zênite, o que implica em grande variação com a estação do ano e o período do dia (Engelsen, 2010; Jablonski e Chaplin, 2009; 2010; Hosseinezhad e Holick, 2013; D’Orazio, *et al.*, 2013; Greaves, 2014). Os raios UVC (280-100 nm) geralmente são integralmente absorvidos pela camada de ozônio e não penetram na terra (Jablonski, 2004; Scherer e Kumar, 2010; D’Orazio, *et al.*, 2013; Holick, 2016).

A ideia de que o fator seletivo primário para adaptação da cor de pele é a radiação, é baseada no fato de que esse traço fenotípico é mais fortemente correlacionado com radiação ultravioleta do que com qualquer outro fator ambiental (Walter, 1971; Jablonski e Chaplin, 2000; Chaplin, 2004). Um estudo de George Chaplin (2004) mostrou que aproximadamente 86% ($r = 0,927$ $p < 0,0001$) da variação na cor de pele humana pode ser explicada exclusivamente pelos níveis de UVMED no outono. UVMED é uma forma de medir os efeitos da radiação na pele, e significa “dose mínima de eritema”. Em outras palavras, é a quantidade de radiação necessária para produzir um leve rubor em pessoas com pele clara (tipo 1 na escala de Fitzpatrick), sendo principalmente determinado por UVB (Jablonski e Chaplin, 2000; 2010; 2013).

A incidência de radiação é maior em latitudes baixas e altitudes elevadas, portanto povos autóctones de regiões com essas características tendem a ter pele mais escura do que populações originárias de outras regiões (Jablonski e Chaplin, 2000, 2010) (Figura1) (Imagem adaptada de Chaplin, 2004).



Source: Chaplin G.®, *Geographic Distribution of Environmental Factors Influencing Human Skin Coloration*, *American Journal of Physical Anthropology* 125:292–302, 2004; map updated in 2007.

Figura 1: Mapa de distribuição de cor de pele de populações nativas, quantificada por refletância, previsto por modelo de regressão múltipla (Imagem copiada de: <https://www.vitamindwiki.com/Skin+color+and+vitamin+D+%E2%80%93+Jablonski+%E2%80%93+Spring+2012>).

Esses resultados de regressão da cor de pele pela radiação UV ilustram a importância da radiação como agente seletivo. Para elucidar a razão fisiológica que gerou essa forte relação causal é preciso compreender os efeitos para o organismo da penetração da radiação na pele.

1.2 Pressão seletiva em ambientes com alta incidência de radiação: Fotólise da vitamina B9 e câncer de pele

Não havendo quantidades suficientes de eumelanina na pele para absorver e dispersar a radiação, os raios UVA são facilmente capazes de penetrar até a derme (D’Orazio *et al.*, 2013). Nesta camada a radiação tem acesso aos vasos sanguíneos, onde circulam compostos bioativos da vitamina B9, genericamente chamados de folatos. Essas moléculas, em especial o 5-metiltetraidrofolato (5-MTHF) que é a principal forma circulatória do folato no corpo humano, são essenciais para a saúde. Os compostos químicos que compõem a família da vitamina B9 estão ligados às reações de metilação e são importantes na síntese e reparo do DNA. Baixos níveis destas substâncias estão associados a diversos problemas de saúde, como defeitos no tubo

neural, anemia megaloblástica, doenças cardiovasculares, câncer e distúrbios cognitivos (Borradale e Kimlin, 2012; Jablonski e Chaplin, 2013; Scaglione e Panzavolta, 2014; Lucock, *et al.*, 2015).

Os raios UVA são capazes de degradar de forma direta o ácido fólico e indiretamente, por meio da geração de espécies reativas de oxigênio, o 5-MTHF (para mais detalhes sobre esses processos, consultar Borradale e Kimlin, 2012). Baixas quantidades dessas substâncias afetam o sucesso reprodutivo masculino, uma vez que prejudicam a espermatogênese (Mathur *et al.*, 1977; Boxmeer *et al.*, 2009). A deficiência de folato também compromete o desenvolvimento embrionário, gerando defeitos no tubo neural, como espinha bífida, o que se constitui como importante fator seletivo para aumento de pigmentação de pele em ambientes de alta exposição à radiação ultravioleta (Jablonski e Chaplin, 2010).

Ambientes de alta incidência de radiação, principalmente do espectro UVB, também estão associados ao aumento de incidência de câncer de pele, especialmente em indivíduos com pele clara (Scherer e Kumar, 2010; Greaves, 2014). Os cânceres de pele, induzidos pela ação mutagênica da radiação, são classificados em duas categorias principais: os melanomas, originados por alterações nos melanócitos, e os cânceres de pele não-melanoma, originados em outras células epidérmicas, predominantemente nos queratinócitos (Scherer e Kumar, 2010). Alguns estudos (Blum, 1961; Elias e Williams, 2013) argumentam que o câncer de pele não poderia constituir-se como força seletiva relevante para o escurecimento de pele, dado que costuma ocorrer depois da idade reprodutiva e está associado a baixa letalidade.

Outros autores (Robins, 1991; Lewis *et al.*, 2013; Greaves, 2014), no entanto, discordam dessa visão. Alegam que não é adequado usar parâmetros atuais de incidência e letalidade da doença para pensar sobre a diminuição no valor adaptativo gerada pelo câncer de pele em outros momentos da história evolutiva humana. Seria preciso contextualizar o papel que esta doença teria tido durante a ocorrência do fenômeno. O processo mais evidente de alteração da pigmentação na linhagem Homininea teria ocorrido quando populações com a pele clara habitavam ambientes com mais incidência de radiação UV e teriam menos recursos culturais para prevenir a doença, tais como roupas, protetores

solares, além de menos conhecimento sobre as implicações da exposição intensiva ao sol. Desse modo, passariam mais tempo ao ar livre e não haveria prevenção nem tratamento de saúde apropriado. Além disso, o câncer do tipo melanoma, embora seja bastante mais raro que os demais, ainda nos dias de hoje acomete jovens em idade reprodutiva, e muitas vezes leva a óbito. Então ao menos esse tipo de câncer é reconhecido como tendo influência seletiva, ainda que moderada, para o aumento de pigmentação (Jablonski e Chaplin 2010).

1.3 Pressão seletiva em ambientes com baixa incidência de radiação: Fotossíntese de vitamina D

A vitamina D, é na verdade um pró-hormônio do hormônio esteroide 1 α ,25-dihidroxitamina D, também chamado de calcitriol (Pilz, *et al.*, 2016). Trata-se de um elemento fundamental para manutenção da homeostase do cálcio e está envolvido em processos fisiológicos essenciais como mineralização óssea. Baixas quantidades dessa substância são a causa de problemas esqueléticos como raquitismo, estreitamento da pélvis e osteoporose (Jablonski e Chaplin, 2009; Veldurthy, *et al.*, 2016). Além dessas ações clássicas, o calcitriol também gera outros efeitos biológicos importantes. Ele está associado, por exemplo, à inibição da progressão de vários tipos de câncer, protege de muitas doenças autoimunes e exerce efeitos sobre o sistema cardiovascular (Christakos, *et al.*, 2016).

A vitamina D pode ser obtida por meio de duas fontes: a alimentação, que provê cerca de 20% da quantidade requerida pelo organismo; e a síntese cutânea, responsável por 80% (Pilz, *et al.*, 2016). Além do nível de exposição à radiação, a produção cutânea de vitamina D depende da pigmentação de pele já que quanto maior a quantidade de melanina menor a penetração de radiação. Peles escuras produzem até seis vezes menos vitamina D que peles claras (Juzeniene *et al.*, 2009; Engelsen, 2010). A produção endógena do hormônio da vitamina D ocorre a partir de um conjunto de reações fotoquímicas e enzimáticas e envolve uma série de genes, conforme será descrito a seguir.

A penetração da radiação ultravioleta do espectro B na pele, especialmente nos comprimentos de onda entre 295 nm e 297 nm, provoca a quebra do 7-deidrocolesterol (7-DHC), que é um esteroide localizado na membrana plasmática dos queratinócitos. Essa reação fotolítica gera a pré-vitamina D₃, que sofre um rearranjo sensível à temperatura, convertendo-se em vitamina D₃, o colecalciferol. Exposição continuada ao sol pode converter a previtamina D₃ em fotoprodutos inativos, o taquisterol e o lumisterol, evitando intoxicação de vitamina D induzida pelo sol (Hossein-nezhad e Holick, 2013; Lucock, *et al.*, 2015; Christakos, *et al.*, 2016; Veldurthy, *et al.*, 2016).

Uma vez formada, a vitamina D₃ é ejetada da membrana plasmática dos queratinócitos e levada até o fígado pela proteína de ligação à vitamina D, DBP (codificada pelo gene *GC*). No fígado o colecalciferol sofre uma hidroxilação no carbono 25, catalisada pela enzima 25-hidroxilase (codificada pelo gene *CYP2R1*), transformando-se em calcidiol, 25(OH)D₃. O calcidiol é transportado pela DBP até os rins, onde as proteínas receptoras transmembrana megalina (codificada pelo gene *LRP2*) e cubilina (codificada pelo gene *CUBN*), localizadas nos túbulos proximais, ligam-se a DBP e promovem a internalização do 25(OH)D₃ (Hossein-nezhad e Holick, 2013; Jolliffe, *et al.*, 2015; Veldurthy, *et al.*, 2016). No interior das células do túbulo proximal, o calcidiol sofre outra hidroxilação, desta vez no carbono 1 por meio da enzima 1 α -hidroxilase codificada pelo gene *CYP27B1*, convertendo-se no hormônio secosteroide calcitriol, 1 α ,25-dihidroxitamina D₃ (1,25[OH]₂D₃) (Engelsen, 2010; Hossein-nezhad e Holick, 2013; Christakos, *et al.*, 2016; Veldurthy, *et al.*, 2016).

O calcitriol tem o efeito de reter o cálcio no organismo, ele provoca aumento da absorção intestinal desse elemento, melhora a reabsorção renal e favorece a mineralização óssea. A atividade da enzima 1 α -hidroxilase, que o produz, é estreitamente regulada tendo sua expressão aumentada pelo paratormônio (PTH), hipocalcemia e hipofosfatemia e é inibida pela hiperfosfatemia, pelo fator 23 de crescimento do fibroblasto (FGF23) e pelo próprio 1,25(OH)₂D₃ (Kini e Nandeesh, 2012; Hossein-nezhad e Holick, 2013; Veldurthy, *et al.*, 2016).

Em ocasiões de excesso de cálcio e vitamina D no organismo, pode ocorrer a expressão do gene *CYP24A1*, que codifica a enzima que hidroxila o carbono 24 tanto do calcidiol quanto do calcitriol, gerando respectivamente os produtos $24,25(\text{OH})_2\text{D}_3$ e $1,24,25(\text{OH})_3\text{D}_3$ que são metabolizados a ácido calcitriólico e excretados (Christakos, *et al.*, 2016; Pilz, *et al.*, 2016; Holick, 2016).

O calcitriol também desempenha um papel importante na regulação da expressão gênica. As ações genômicas do calcitriol são mediadas pelo fator de transcrição VDR (codificado por um gene de mesmo nome). O VDR é membro da superfamília de receptores nucleares de hormônios que regulam a transcrição, ele possui mais de 900 polimorfismos de nucleotídeo único (SNPs) em populações humanas, alguns deles associados a fenótipos clínicos. Os SNPs mais frequentemente estudados neste gene são rs1544410, rs731236, rs2228570 e rs7975232, conhecidos pelo nome das enzimas de restrição que os clivam, respectivamente *BsmI*, *TaqI*, *FokI* e *Apal* (Santoro *et al.*, 2015; Nunes *et al.*, 2016; Alizadeh *et al.*, 2017).

Ao ligar-se a $1,25(\text{OH})_2\text{D}_3$ o VDR entra no núcleo, onde sofre uma heterodimerização, associando-se ao receptor retinoide X (RXR), codificado pelo gene *RXRA*. O complexo ativo $1\alpha,25(\text{OH})_2\text{D}_3$ -VDR-RXR, interage com regiões específicas no DNA, os elementos de resposta à vitamina D (VDREs) localizados em genes alvo ou em seu entorno. Essa interação leva ao recrutamento de outras proteínas que modulam a expressão gênica resultando na ativação ou repressão da transcrição. O VDR é quase ubíquo, está presente em mais de 38 tecidos e diversos tipos celulares, dentre eles o epitélio intestinal, mamário, túbulos renais, glândula pituitária, paratireoide, tecidos reprodutivos, queratinócitos, células do sistema imune, etc. (Haussler *et al.*, 2011; Lucock, *et al.*, 2015). Estima-se que o calcitriol regule, direta ou indiretamente, 2.000 genes que exercem diversas funções no genoma humano, indicando a importância desse hormônio (Hosseini-nezhad e Holick, 2013; Christakos, *et al.*, 2016; Holick, 2016). O esquema apresentado na Figura 2 ilustra os processos supramencionados.

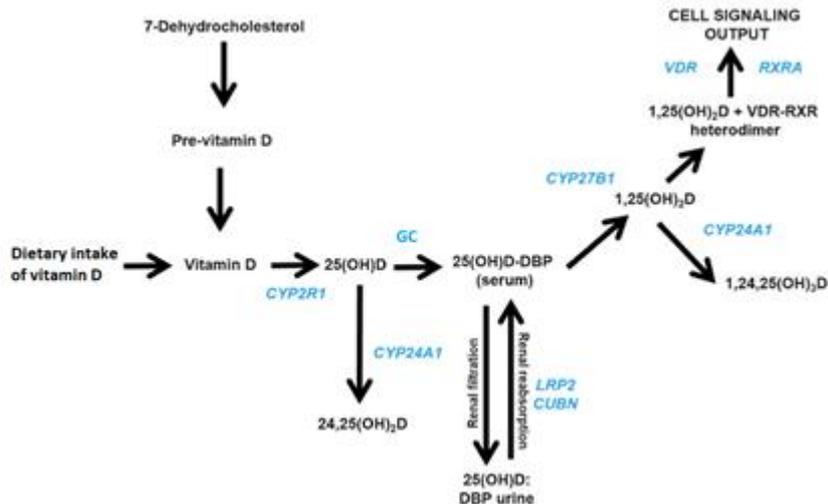


Figura 2: Diagrama que descreve as etapas das vias de metabolismo e sinalização da vitamina D. Em azul estão os principais genes envolvidos no metabolismo e controle da expressão, cujas funções estão descritas no texto. Este esquema é uma adaptação de Jolliffe, *et al.* (2015) e contém a lista de genes, relacionados a vitamina D, candidatos deste estudo.

A importância de produção cutânea e os efeitos deletérios da falta dessa substância, levaram à formulação da hipótese de que ambientes com pouca incidência de radiação seriam vantajosos seletivamente para indivíduos com pele clara. Isso constitui-se na 'hipótese da vitamina D', que é a teoria mais aceita atualmente para explicar por que populações autóctones de regiões com baixa incidência de radiação tendem a ter a pele clara.

1.4 Evolução da cor de pele humana e a adaptação das populações nativas continentais à incidência de radiação

Acredita-se que o ancestral comum aos humanos e chimpanzés tinha pele clara coberta de pelos, à semelhança da maioria dos atuais primatas pertencentes à infraordem Catarrhine (Jablonski e Chaplin 2000; 2010; Jablonski, 2004). Após a divergência entre as duas linhagens, estimada aproximadamente 6,5-9,3 milhões de anos atrás (Moorjani *et al.*, 2016), houve seleção para perda de pelos corporais no grupo que deu origem aos humanos. Não é possível datar o surgimento desse traço, pois não são conhecidas as mutações subjacentes a ele (Stoneking, 2017). Alguns trabalhos argumentam

que a perda de pelos corporais teria surgido em resposta à necessidade aumentada de termorregulação. Segundo essa hipótese, o evento teria começado a ocorrer progressivamente há 2 milhões de anos, devido ao estresse térmico gerado pela migração para ambientes abertos de Savana e adoção de atividades musculares extenuantes, como a caça diurna (Wheeler, 1992; Rogers *et al.*, 2004; Ruxton e Wilkinson, 2011).

De qualquer modo, a perda de pelos corporais levou à alteração em uma série de estruturas da pele, selecionando variantes em genes responsáveis pela diminuição da permeabilidade e aumento da resistência à abrasão (Jablonski e Chaplin, 2010). Este fenótipo também causou diminuição da proteção contra radiação ultravioleta (RUV), que degrada o folato, vitamina importante na síntese e no reparo do DNA que protege contra defeitos congênitos na formação do tubo neural. Isso resultou em pressão seletiva para aumento da produção cutânea do pigmento melanina, que é fotoprotetivo, e consequente escurecimento da pele ao longo das gerações (Jablonski e Chaplin, 2000, 2010, 2017). Há cerca de 1,2 milhões de anos a seleção favoreceu uma variante do gene *MC1R*, receptor de melanocortina 1, que confere pele escura às populações africanas e está sob seleção purificadora neste continente (Rogers *et al.*, 2004). Análises mostraram sinais de seleção positiva para esse alelo, indicando também que o processo de escurecimento da pele provavelmente começou a ocorrer antes do surgimento dos humanos modernos, datado entre 200 mil e 300 mil anos atrás (Stoneking, 2017; Deng e Xu, 2018; Crawford, *et al.*, 2017). Além de *MC1R*, variantes em outros genes da via da pigmentação também estariam segregando por centenas de milhares de anos na linhagem dos hominíneos (Crawford, *et al.*, 2017).

Quando algumas populações humanas modernas saíram da África, cerca de 100 mil anos atrás em direção a latitudes norte, passaram a viver em um ambiente com menos exposição à radiação. Como já mencionado, a radiação UVB permite a fotossíntese cutânea de vitamina D, que embora possa ser adquirida através da dieta, tem a produção endógena como principal fonte. A menor incidência de UVB leva à diminuição deste pró-hormônio, cuja deficiência é causa de uma série de problemas de saúde, incluindo alguns diretamente relacionados à diminuição do valor adaptativo, como estreitamento

da pélvis, que gera problemas no parto, e raquitismo (Jablonski e Chaplin, 2009). Nesse novo ambiente, indivíduos com quantidades menores de melanina na pele, e, portanto, mais facilidade na etapa inicial de síntese da vitamina D, passaram a ter vantagem seletiva.

No Oriente Médio, há aproximadamente 60-50 mil anos, os humanos anatomicamente modernos entraram em contato com os Neandertais (*Homo neanderthalensis*, espécie de hominíneo arcaico que divergiu do *Homo sapiens* moderno há ~430 mil anos e viveu na Eurasia entre 400-28 mil anos atrás) (Deng e Xu, 2018; Kuhlwilm, *et al.*, 2016). Não se sabe ao certo qual era a cor de pele desses hominíneos, entretanto a maioria dos autores acredita que eles possuíam a pele clara, quando comparada aos africanos do sul do Saara (Juzeniene *et al.*, 2009; Stoneking, 2017; Deng e Xu, 2018; Crawford, *et al.*, 2017).

De forma pioneira, nosso grupo de pesquisa, utilizando 124 SNPs localizados em 33 genes e 7 regiões intergênicas ligados a via da pigmentação, desenvolveu um método para predição dos fenótipos de cor da pele, cabelo e olhos a partir de genomas completos já conhecidos (Cerqueira *et al.*, 2012). Os resultados indicaram que os 3 Neandertais amostrados na atual Croácia (Vi33.16, Vi33.25 e Vi33.26) teriam uma pele relativamente escura (similar a dos atuais europeus do mediterrâneo) e cabelo castanho, exceto o indivíduo Vi33.25 que seria ruivo. Anteriormente, Lalueza-Fox *et al.* (2007) encontraram uma variante no gene *MC1R* em dois Neandertais que habitavam a atual Espanha, similar a que nos humanos modernos determinaria pele clara e cabelos ruivos.

A partir do conhecimento do primeiro genoma nuclear completo do Neandertal (Green e cols. 2010), foi possível também identificar a dimensão do intercruzamento entre as duas espécies. Cerca de 2% do genoma das populações não-africanas modernas seria de origem Neandertal (Skoglund e Mathieson, 2018). Como resultado, alguns autores postulam que houve a entrada de alelos Neandertais relacionados à pigmentação mais clara em algumas populações da nossa espécie, novamente considerando uma comparação com africanos do sul do Saara. Esses alelos teriam sido selecionados positivamente, conforme sugerido a partir de dados que apontam

que nas populações europeias, um haplótipo que contém um alelo de origem Neandertal no gene *BNC2*, têm frequência de 70%. Esse haplótipo codifica uma proteína com função relacionada a saturação da cor da pele (Jacobs *et al.*, 2013; Marciniak e Perry, 2017).

Outros autores demonstraram ainda uma origem de variantes ligados ao clareamento da pele na nossa espécie. Por exemplo, um alelo a montante do gene *KITLG* (que codifica uma proteína com amplo espectro de atuação, incluindo melanogênese) associado à pele mais clara teria sido selecionado há cerca de 30 mil anos no ancestral comum das populações asiáticas e europeias (Beleza, *et al.*, 2012; Deng e Xu, 2018; Stoneking, 2017). Dados do HGDP (*Human Genome Diversity Project*) por sua vez mostraram que essa variante de *KITLG* também teria se espalhado pela Oceania, e é bastante frequente nas populações ameríndias estudadas até o presente (Deng e Xu, 2018).

Em contrapartida, um número significativo de variantes populacionais-específicas ligadas a vários outros genes da via da pigmentação, e que estariam associados ao desenvolvimento do fenótipo “pele clara”, teriam emergido de forma independente em populações da Ásia e Europa. Esse fato indica convergência adaptativa e denota a importância da seleção natural na evolução da pigmentação cutânea humana. Por exemplo, a partir de dados obtidos de populações europeias atuais, alelos associados à cor clara nos genes *SLC24A5*, *SLC45A2* (identificados como genes determinantes na pigmentação humana e de outros vertebrados) e *TYRP1* (envolvido na produção de melanina) tiveram seleção estimada entre 19 e 10 mil anos atrás (Beleza, *et al.*, 2012). Estima-se que o alelo relacionado à cor clara no gene *SLC45A2*, embora presente desde 9 mil anos AP na Escandinávia, apresentava frequências muito baixas ou estava ausente na maior parte da Europa antes do início da Era do Cobre (~5.300- 4.800 AP). As variantes associadas à cor de pele clara nos genes *TYR* e *HERC2*, também aumentaram sua frequência mais recentemente, entre 6.500 e 4.000 anos AP na atual Ucrânia (Marciniak e Perry, 2017).

Populações do leste da Ásia têm um alelo relacionado à cor clara no gene *OCA2* que só existe nesta região, então é suposto que se trate de uma

mutação nova que foi selecionada neste local. A exata função da proteína codificada por *OCA2* é desconhecida, mas sabe-se que é essencial para a pigmentação normal e provavelmente está envolvida na produção de melanina. Asiáticos do leste também mostraram seleção positiva exclusiva em uma variante no gene *DCT*, um importante parálogo de *TYRP1* (Deng e Xu, 2018). O gene *MFSD12*, que influencia como a melanina é produzida e armazenada na pele, tem um alelo derivado no leste da Ásia com frequências correlacionadas com baixa intensidade da radiação solar. Esse alelo está ausente na Europa e é relacionado a cor clara, indicando mais uma vez a evolução paralela da despigmentação na Europa e Ásia (Adhikari *et al.*, 2019).

Como um todo, essas investigações com DNA de populações antigas revelaram que o processo de despigmentação possivelmente ocorreu paulatinamente e se acentuou mais recentemente na Eurásia. Outra informação interessante advinda de estudos com DNA de amostras arqueológicas é a possibilidade de que a despigmentação da pele tenha sido influenciada também pela agricultura e não apenas pela incidência de radiação. Um alelo relacionado à pele clara no gene *SLC24A5* estava ausente ou existia em baixas frequências em caçadores-coletores que viveram entre 13 mil e 7 mil anos AP na Europa. Porém, tal alelo seria encontrado em altas frequências ou fixado em agricultores do neolítico que habitavam a Anatólia, região que hoje compõe a maior parte da Turquia, e chegaram na Europa por volta de 7,5 mil anos AP. Este fato sugere que os migrantes desta região levaram tal alelo para a Europa ocidental, sendo o mesmo hoje encontrado em altas frequências em todo continente (Mathieson *et al.* 2015; Marciniak e Perry, 2017).

Um aspecto importante que deve ser mencionado é que uma dieta baseada em plantas domesticadas, comuns em sociedades agriculturalistas, normalmente tem baixas quantidades de vitamina D, em contraste com a alimentação de caçadores-coletores, mais rica e diversa pois contém produtos de caça e recursos marinhos. As evidências arqueológicas de deficiência de vitamina D, baseadas em deformidades esqueléticas e defeitos na mineralização dentária, são raras antes das origens da agricultura e aumentaram ao longo do tempo à medida que as populações humanas deixaram os hábitos de caça e coleta e tornaram-se agriculturalistas e

sedentárias (Brickley *et al.*, 2017). Então, no período de expansão da agricultura, sedentarização e, conseqüentemente, aumento populacional, a seleção de alelos que conferem pele mais clara pode ter sido impulsionada como adaptação para compensar a baixa ingestão da vitamina D e suas conseqüências deletérias (Richards, *et al.*, 2003; Marciniak e Perry, 2017; Stoneking, 2017; Quillen *et al.* 2019).

Como pode ser visto até aqui, os estudos sobre a seleção de alelos de genes relacionados com pigmentação estão predominantemente concentrados em populações europeias e asiáticas, mas alguns dados sobre outros grupos continentais são reveladores. Populações da Oceania, por exemplo, apresentam em altas frequências uma variante específica do gene *OCA2* (Sturm e Duffy, 2012). Há também uma variação não sinônima exclusiva no gene *TYRP1* em habitantes das Ilhas Salomão e em algumas partes do arquipélago de Bismarck. Essa variante está relacionada ao cabelo claro e também, de modo menos acentuado, à cor de pele (Sturm e Duffy, 2012; Deng e Xu, 2018). Foram encontradas variantes em torno de diversos genes, tais como *SLC24A5*, *MFSD12*, *DDB1*, *TMEM138*, *OCA2*, *HERC2*, *SLC24A5*, *TYRP1*, *SMARCA2/VLDLR*, *SNX13*, associadas a diferenças de cor em populações africanas (Martin, *et al.*, 2017; Crawford *et al.*, 2017).

Esses achados como um todo indicam heterogeneidade e complexidade na arquitetura genética por trás da pigmentação humana e mostra a importância de estudar diferentes populações.

1.5 Variação da cor da pele em Nativos Americanos: variabilidade genética e implicações adaptativas

Pouco se sabe sobre pigmentação de pele em nativos americanos. Essas populações têm origem em um grupo de indivíduos que começou a divergir dos asiáticos em torno de 36 mil anos atrás, ainda no nordeste da Ásia. Recentes estudos indicam ainda que houve fluxo gênico com outros asiáticos e com povos norte-eurasianos até aproximadamente 25 e 20 mil anos AP, respectivamente (Moreno-Mayar *et al.*, 2018a).

O resfriamento da Terra durante o último máximo glacial (entre 26 e 18 mil anos AP) ocasionou a diminuição do nível do mar e consequente exposição da Beringia, uma grande massa de terra que juntava o nordeste da Ásia ao noroeste da América. Este local, onde ingressaram os ancestrais dos nativos americanos entre 24 e 20 mil anos AP, era então caracterizado por um clima relativamente ameno (González-José *et al.*, 2008; Fagundes *et al.*, 2008; Bortolini *et al.*, 2014; Amorim, *et al.*, 2017; Moreno-Mayar *et al.*, 2018a; Pinotti, *et al.*, 2019). Não há consenso sobre o tempo em que a Beringia foi habitada, as estimativas variam entre 2,4 mil até 8 mil anos (Raghavan *al.*, 2015; Amorim, *et al.*, 2017; Pinotti, *et al.*, 2019), um tempo estimado como suficiente para a emergência de alelos exclusivos, só encontrados em populações nativas americanas (Amorim, *et al.*, 2017; Posth, *et al.*, 2018).

Entre 18 mil anos AP e o fim do pleistoceno (~12-10 mil anos AP), a massa de terra da Beringia foi sendo reduzida devido ao ótimo climático e, então, aumento do nível do mar. A diminuição de recursos, somada ao aumento populacional, impulsionou a migração de algumas das populações que viviam na Beringia para o sul. Como resultado, entre 19,5 e 15 mil anos AP houve a efetiva entrada no continente americano (González-José *et al.*, 2008; Fagundes *et al.*, 2008; Reich *et al.*, 2012; Bortolini *et al.*, 2014; Amorim, *et al.*, 2017; Pinotti, *et al.*, 2019).

Ao chegarem na América, os primeiros migrantes se depararam com uma ampla variedade de ambientes onde desenvolveram diferentes culturas. Em algumas regiões, como nos Andes e na Mesoamérica, as sociedades se tornaram urbanas e produtoras de agricultura, outras comunidades permaneceram caçadoras-coletoras até os dias atuais (Hünemeier *et al.*, 2012; Park *et al.*, 2016; Reales *et al.*, 2017).

Vale destacar que a América abriga populações autóctones que habitam regiões com extrema diferença na incidência de radiação UV, desde as altas latitudes do Ártico até elevadas altitudes dos Andes. Pesquisas sobre a cor de pele atual de índios da América são escassas. As informações diretamente observadas mais completas que dispomos são apresentados em um mapa na escala de von Luschan, com dados obtidos antes de 1940 pelo geógrafo italiano Renato Biasutti (figura 3) (Barsh, 2003; Jablonski, 2004; Lewis

et al., 2013). Esse mapa é impreciso, uma vez que o método de medição é subjetivo e as regiões para as quais não havia informações o autor preencheu por extrapolação com os resultados de outras áreas (Robins, 1991; Jablonski, 2004).

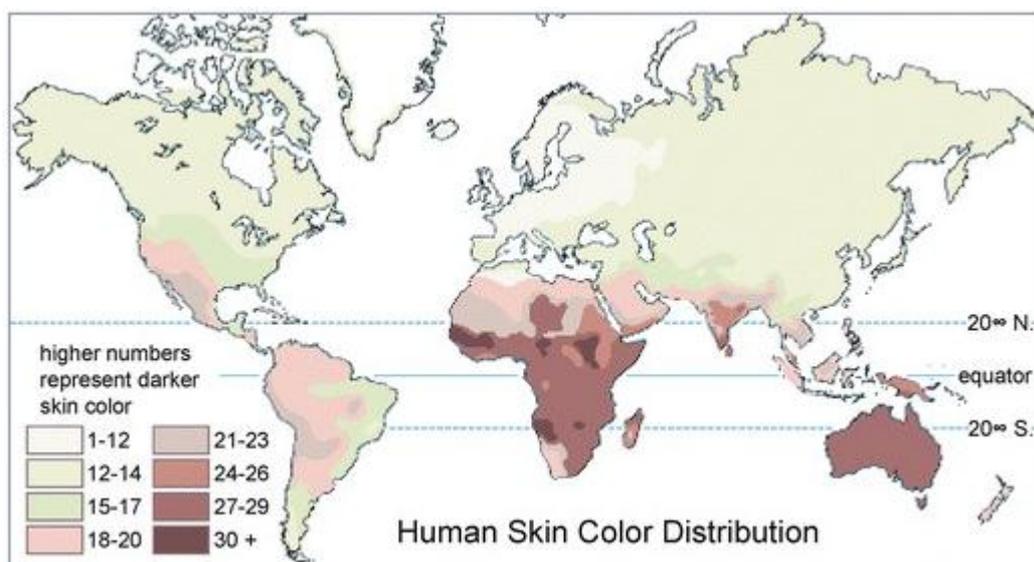


Figura 3: Mapa de distribuição de cor de pele em populações humanas na escala de von Luschan (copiado de Deng e Xu, 2018).

Dados mais precisos sobre cor de pele de nativos americanos, com medida de refletância, são apresentados em poucos trabalhos. Jablonski e Chaplin (2000) compilaram as informações então disponíveis para todas as populações autóctones do mundo. Os autores mostram que existem dados de Inuits (esquimós) do oeste da Groenlândia e de quatro populações da América do sul.

Apesar da escassez de trabalhos sobre o tema nesse conjunto de populações, os indígenas americanos são frequentemente citados (Murray, 1934; Diamond, 1991; Robins, 1991; Aoki, 2002; Barsh, 2003; Jablonski, 2004; Juzeniene *et al.*, 2009; Jablonski e Chaplin, 2000; 2010; 2013; Elias e Williams, 2013; Greaves, 2014; Quillen, 2015; Deng e Xu, 2018; Adhikari *et al.*, 2019; Quillen *et al.* 2019) como exceções à regra de distribuição da cor de pele de populações autóctones. De fato, se compararmos a cor de pele esperada para nativos do continente americano apresentada na figura 1 com as informações obtidas de observação direta destas populações, mostradas na figura 3, percebe-se uma grande diferença. Notadamente os Inuit têm a pele mais

escura do que seria esperado para as latitudes em que eles habitam, e consequente incidência de radiação a que estão expostos. Por outro lado, as populações nativas americanas de latitudes equatoriais e os habitantes das elevadas altitudes andinas têm a pele mais clara do que o esperado.

A inexistência de um gradiente latitudinal de cor de pele equivalente ao encontrado em populações do Velho Mundo pode ter uma base genética, conforme indicado por Adhikari *et al.*, (2019). Ao analisar um amplo conjunto de dados de 6.000 latino-americanos, os autores não detectaram nenhuma variante relacionada a cor de pele exclusiva da América, porém descreveram novas regiões associadas com a pigmentação da pele (em 10q26) e cor do olho (em 1q32, 20q13 e 22q12). Além disso, detectaram uma forte associação com a pigmentação da pele e uma variante (Tir182His; na proteína do gene *MFSD12*) comum, como já mencionado, apenas em asiáticos e nativos americanos. Os autores mostraram que a frequência do alelo derivado (182His) está significativamente correlacionada com a menor intensidade de radiação solar no leste da Ásia, indicando que *MFSD12* esteve sob pressão seletiva em asiáticos, provavelmente após sua separação dos europeus (Adhikari *et al.*, 2019).

Outro estudo que merece destaque é o de Quillen e colaboradores (2012). Neste artigo os autores mostraram que, diferentemente do encontrado em outras populações, um conjunto dos genes de pigmentação em indígenas americanos não teria sinal de ter sofrido ação de seleção, quando comparado com um conjunto aleatório de outros genes do genoma. Todavia isso não significa que determinados genes de pigmentação não tenham sofrido seleção nessas populações, conforme mostrou o mesmo trabalho ao encontrar variantes dos genes *OPRM1* e *EGFR* com sinal da ação de seleção, além de associação com diferenças na cor de pele nos grupos sob investigação (Quillen *et al.*, 2012).

Alguns dos trabalhos que criticam a teoria segundo a qual a cor de pele é uma adaptação à incidência de radiação (Diamond, 1991; Aoki, 2002), utilizam essa discrepância fenotípica para respaldar a idéia de que a pigmentação cutânea não é explicada pela seleção natural. Os autores que defendem que a cor de pele é resultado da seleção natural acreditam que a

inadequação da cor de pele dos nativos americanos pode ser devido a sua chegada relativamente recente ao continente (Jablonski, 2004; Juzeniene *et al.*, 2009; Quillen, 2015; Adhikari *et al.*, 2019). Portanto, segundo eles, essas populações poderiam não ter tido tempo suficiente para se adaptar aos ambientes em que habitam. Com relação aos Inuit especificamente, a alimentação tradicional baseada em mamíferos marinhos e peixes, que é rica em vitamina D, também é usada como justificativa para um padrão de cor diferente daquele esperado (Murray, 1934; Quillen *et al.* 2019). Sobre os indígenas que tem a cor mais clara do que o esperado, além da proposta da colonização recente, são consideradas também outras hipóteses para explicar o fenômeno. Por exemplo, que o uso de roupas poderia ser um fator atenuante para pressão seletiva (Jablonski, 2004), ou que os americanos poderiam ter capacidade aumentada de produzir pigmentação facultativa (bronzamento) (Quillen, 2015).

De fato, o continente americano foi o último a ser habitado por populações humanas e isso poderia constituir-se como fator atenuante para a adaptação. Contudo, trabalhos de nosso grupo de pesquisa e de outros autores têm mostrado que as populações nativas americanas, apesar do relativo pouco tempo no continente, sofreram adaptação específicas, por exemplo: às elevadas altitudes andinas (Jacovas, *et al.*, 2015; Jacovas, *et al.*, 2018), ao ambiente glacial e à dieta protéica ainda na Beringia (Amorim, *et al.*, 2017), às regiões de floresta tropical (Amorim, *et al.*, 2015) e à agricultura desenvolvida na Mesoamérica (Hünemeier *et al.*, 2012). Além disso, conforme exemplo de populações européias, em que parte significativa da despigmentação foi relativamente rápida e recente (entre 12 e 6 mil anos AP), não são exigidas necessariamente muitas gerações para a adaptação deste fenótipo (Gibbons, 2007). Contudo, é preciso ponderar que não se espera resposta molecular exatamente igual considerando diferentes grupos continentais em contextos de evolução paralela, embora estejam sujeitos a similares pressões seletivas e que as mesmas vias sejam cooptadas, esses grupos apresentam um relativamente diferente repertório genético (incluindo nas vias que podem “responder” a determinada pressão seletiva), além de terem passado por diferentes eventos estocásticos e demográficos.

Um ponto central neste debate que é muitas vezes ignorado, embora esteja subjacente à teoria, é que a cor de pele sofre seleção atrelada à resposta fisiológica à radiação UV. Ou seja, o fenótipo que está sendo selecionado é determinado por um amplo conjunto de genes, em vias diferentes, relacionadas tanto à pigmentação quanto ao metabolismo e controle da expressão da vitamina D. Desse modo, medir apenas a cor de pele para determinar se houve adaptação é limitado. Trabalhos de vários autores (Hochberg e Templeton, 2010; Lucock, *et al.*, 2014; 2015; Tiosano, *et al.*, 2016) sustentam a importância do estudo da adaptação à radiação de maneira integrada, considerando a variação em genes pertencentes às diferentes vias que com potencial para interferir no fenótipo.

Hochberg e Templeton (2010) propuseram que além dos genes relacionados à pigmentação, variantes alélicas no gene *VDR* (controle da expressão da vitamina D) integram o complexo evolutivo que adapta os seres humanos a diferentes níveis de exposição à radiação UV. Lucock e colaboradores (2015) ao revisar 28 estudos, encontraram uma relação entre a frequência de alelos específicos de 4 polimorfismos conhecidos no gene *VDR* (*BsmI* (b), *Apal* (a), *TaqI* (t) e *FokI* (f)) e a latitude em populações africanas e eurasiáticas localizadas entre 12° e 60° N. Com base em um trabalho anterior conduzido por Lucock (2014), os autores ainda propuseram hipóteses sobre os mecanismos moleculares por meio dos quais a radiação UV selecionaria variantes específicas do *VDR* e de genes centrais no metabolismo do folato. Isso constitui-se como um possível artifício evolutivo para adaptação conjunta dessas vitaminas fotossensíveis e a pigmentação.

A partir de um método que busca interação gênica, Tiosano e colaboradores (2016) mostraram a existência de coadaptação entre genes de pigmentação e o *VDR*. Este trabalho encontrou uma relação latitudinal entre conjuntos de polimorfismos de genes relacionados à pigmentação e variantes do *VDR*. Esses polimorfismos, de ambas as vias, encontrados em desequilíbrio de ligação foram agregados no que os autores chamaram de 'redes'. Uma informação central deste trabalho é que alguns dos SNPs pertencentes a essas redes, vistos de modo isolado, não se relacionam com a latitude. Ou seja, apenas dentro de um contexto mais amplo, variantes alélicas funcionais de

diferentes vias são selecionadas em conjunto em resposta aos diferentes níveis de radiação, que se alteram com a latitude. Portanto, foi possível comprovar estatisticamente a existência de epistasia relacionada ao valor adaptativo de uma característica fenotípica humana.

As populações nativas americanas, por suas características peculiares são especialmente interessantes para investigar a existência de adaptação e coadaptação nessas vias. Conforme dito anteriormente, essas populações habitam uma ampla gama de latitudes, possuem diferentes hábitos de subsistência e alimentares, além de apresentarem distribuição da cor de pele em desacordo com o esperado pela distribuição geral dessa característica entre os povos autóctones de outros continentes (os habitantes de latitudes equatoriais, como amazônicos e os andinos têm a pele mais clara e os Inuit têm a pele mais escura do que o esperado).

Vale destacar que o nosso trabalho será o primeiro a investigar, além do *VDR* e outros genes de pigmentação, polimorfismos ligados ao controle da expressão e a via metabólica da vitamina D, bem como a interação entre todos eles, acrescentando assim maior complexidade à investigação. Com isso buscamos identificar possíveis combinações alélicas que teriam um potencial funcional e adaptativo considerando os nativos americanos como um todo, bem como grupos populacionais que habitam diferentes ambientes e/ou que possuem distintos hábitos de subsistência e dieta.

Objetivos:

2.1 Objetivo geral

Avaliar o padrão de diversidade de genes relacionados à pigmentação de pele, metabolismo e controle da expressão da vitamina D em 333 indivíduos pertencentes a 33 populações nativas americanas que habitam diferentes ecossistemas e que possuem diferenças em suas práticas de subsistência/hábitos de dieta. Investigar a ocorrência de adaptação e coadaptação em alelos desses genes.

2.2 Objetivos específicos

- Estimar as frequências alélicas dos SNPs dos genes candidatos em 333 indivíduos de 33 populações Nativas Americanas;
- Determinar se as frequências alélicas destes genes têm associação com variáveis ambientais de interesse, ou seguem o padrão esperado pela teoria neutra de evolução molecular;
- Investigar a presença de associação multi-SNPs em alelos de genes relacionados à via da vitamina D e da pigmentação de pele em populações Nativas Americanas;
- Calcular a frequência das associações multi-SNPs em cada uma das categorias populacionais;
- Verificar se existe relação entre interações alélicas e variáveis ambientais de interesse (altitude, latitude), a fim de inferir seleção conjunta (coadaptação) de alelos de genes dessas duas rotas em resposta a incidência de radiação UV;
- Verificar se existe relação entre as interações alélicas e o modo de subsistência das populações (caçadores-coletores ou agriculturistas) para detectar coadaptação de alelos de genes dessas duas vias em resposta a quantidade de vitamina D na dieta.

Amostra e Métodos

3. Populações e Genes

3.1 Populações amostradas

A amostra é composta por um conjunto de sete bancos de dados, com um total de 364.470 SNPs, conforme descrito em Reich *et al.* (2012). Dentre os indivíduos da varredura genômica mencionada, foram escolhidos nativos americanos da Amazônia, dos Andes, da Mesoamérica e de latitudes extremas, como os esquimós da Groenlândia. Os dados foram curados para excluir indivíduos com altos níveis de miscigenação, resultando em 333 indivíduos pertencentes a 33 populações nativas americanas. Essas populações pertencem a diferentes etnias, oriundas de regiões que compreendem uma ampla gama de latitudes e condições ambientais (Figura 4). Informações adicionais sobre as populações encontram-se na Tabela Suplementar 1 ao final do texto.

Aprovação ética para o uso dessas amostras em estudos evolutivos e antropológicos foi fornecido pela Comissão de Ética Nacional (Resolução CONEP nº123/98). Quando havia ausência de consentimento livre e esclarecido por escrito, consentimento verbal informado individual e tribal foi obtido dos participantes analfabetos de acordo com a Declaração de Helsinki.

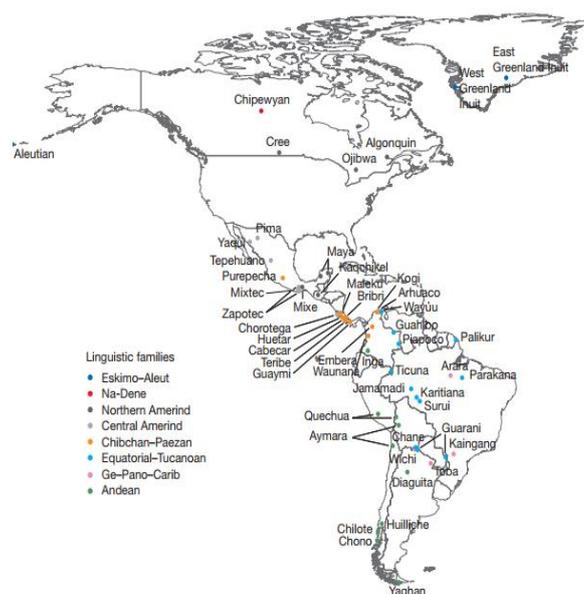


Figura 4- Localização geográfica e famílias linguísticas das populações nativas americanas amostradas, seguida pelo seu tamanho amostral: Arara (n=1), Arhuaco (n=5), Aymara (n=23),

Inuit/esquimó do leste da Groenlândia (n=6), Inuit/esquimó do oeste da Groenlândia (n=4), Guahibo (n=6), Guarani (n=6), Jamamadi (n=1), Kaingang (n=1), Kaqchikel (n=12), Karitiana (n=13), Kogi (n=4), Maleku (n=3), Maya (n=48), Mixe (n=17), Mixtec (n=5), Palikur (n=3), Parakanã (n=1), Piapoco (n=7), Purepecha (n=1), Quechua (n=40), Surui (n=24), Tepehuano (n=25), Teribe (n=3), Ticuna (n=6), Toba (n=4), Waunana (n=3), Wayuu (n=9), Wichi (n=5), Yaghan (n=4), Zapotec (n=43). As populações Maya e Zapotec têm evidência de subestrutura, por isso cada uma delas foi representada como duas subpopulações (conforme tabela S1) (Reich *et al.*, 2012).

3.2 Categorias populacionais

As 33 populações amostradas foram classificadas em categorias que potencialmente interferem na seleção de variantes de genes relacionados ao metabolismo/controle da expressão da vitamina D e pigmentação de pele. As categorias levam em conta as variáveis latitude, altitude e práticas de subsistência e hábitos de dieta das populações que as compõem. São elas “agriculturalistas”, “caçadores-coletores”, “terras altas”, “terras baixas”, “altas latitudes” e “baixas latitudes”, conforme descrito abaixo. Uma categoria adicional, denominada de “agriculturalistas de terras baixas”, também foi considerada.

3.2.1 Para testar a influência das práticas de subsistência e hábitos de dieta

Foram consideradas as seguintes categorias: 1) caçadores-coletores (119 indivíduos, 22 populações) e 2) agriculturalistas (214 indivíduos, 11 populações), 333 indivíduos ao total, conforme tabela suplementar 1. Sabemos que práticas de subsistência e hábitos de dieta mudam com o tempo. Além disso, comunidades de caçadores-coletores também podem praticar algum nível de agricultura incipiente. Contudo, no presente trabalho nós consideramos como agricultores as populações que drasticamente mudaram sua trajetória histórica devido à adoção da agricultura e /ou atividades pastorais na era pré-colombiana. Ou seja, populações que se tornaram sedentárias, promovendo o surgimento do núcleo urbano (por exemplo, andinos e o Império Inca na América do Sul). Por outro lado, consideramos como caçadores-coletores aquelas populações que atualmente (ou até recentemente) seguem estilos de vida similares aos de seus antepassados em as áreas mais remotas da

América do Sul. Basicamente são povos amazônicos, e outros grupos nativos com distribuição mais meridional como os Kaingang e Toba (tabela suplementar 1).

3.2.2 Para testar a influência da altitude:

Foram considerados as seguintes categorias: 1) terras altas (63 indivíduos, 2 populações) e 2) terras baixas (109 indivíduos, 20 populações), totalizando 172 indivíduos.

São classificados como “terras altas” as populações que habitam ambientes com altitudes $\geq 2,500$ m (Aymara e Quechua). Foram escolhidos como representantes da categoria “terras baixas” as outras populações da América do Sul utilizadas neste estudo.

3.2.3 Para discriminar a influência das práticas de subsistência/hábitos de dieta da altitude:

Parte importante da amostra de agriculturistas é composta por populações que vivem em altas altitudes (andinos). Então é possível que a altitude (altas incidências de UV, por exemplo) influencie o resultado da comparação entre as categorias “agriculturistas” e “caçadores-coletores”. Quando são comparadas as categorias “terras altas” e “terras baixas”, por sua vez, é difícil discriminar se o fator que influencia a diferença é a alimentação ou a altitude, já que todos os indivíduos de terras altas são também agriculturistas e todos os de terras baixas são caçadores-coletores. Com objetivo de discernir o papel destas variáveis, criamos uma categoria somente com os agriculturistas de terras baixas (151 indivíduos, 9 populações). Com isso em mente, quando caçadores-coletores são comparados com agriculturistas de terras baixas podemos supor que as diferenças entre estes não podem ser pelo ambiente de altitude, mas sim por outros fatores, como diferenças na alimentação e/ou devido a questões demográficas/estocásticas.

3.2.4 Para testar a influência da latitude:

Foram considerados as seguintes categorias: 1) Altas latitudes (14 indivíduos, 3 populações) e 2) baixas latitudes (314 indivíduos, 28 populações), 328 indivíduos ao total.

Não existe uma definição formal para altas e baixas latitudes. Neste estudo a categoria “altas latitudes” compreende populações que habitam latitudes $\geq |50^\circ|$. As populações classificadas nesta categoria são os esquimós da Groenlândia (Inuit) e os Yaghan, considerados o povo indígena mais meridional do mundo. Baixas latitudes foram definidas considerando a região do equador até os trópicos $\leq |23^\circ 26'|$. Todas as outras populações foram enquadradas nesta categoria, exceto Kaingang (-24°) e Toba (-26.5°), que foram excluídas da análise.

Além da latitude, Yaghan e os povos Inuit tinham em comum, pelo menos até recentemente, uma dieta basicamente composta de mamíferos marinhos (Lowie, 1940).

3.3 Seleção de genes e obtenção de SNPs

Com base na literatura foram selecionados 17 genes candidatos, sendo 8 deles relacionados a vias de metabolismo e controle de expressão da Vitamina D, e outros 9 genes envolvidos com pigmentação de pele.

3.3.1 Metabolismo da vitamina D:

CYP27B1, CUBN, CYP24A1, CYP2R1, LRP2, GC.

3.3.2 Controle da expressão da vitamina D:

RXRA e VDR.

3.3.3 Via da pigmentação da pele:

TYR, TYRP1, OCA2, SLC45A2, SLC24A5, KITLG, MC1R, OPRM1, EGFR.

Informações acerca da função e localização dos genes candidatos desse estudo, bem como dos demais genes citados no trabalho, encontram-se nas tabelas suplementares 2, 3 e 4 ao final do texto.

A partir do banco de dados dbSNP, disponível no NCBI (<https://www.ncbi.nlm.nih.gov/SNP/>), foi compilada uma lista de todos os SNPs já descritos para os 17 genes candidatos. Esta lista resultou em 219.463 SNPs. Destes, encontramos sobreposição com o nosso banco de dados para 15 dos genes candidatos e 408 SNPs. O gene MC1R não tem nenhum polimorfismo genotipado na amostra. Informações sobre o número de SNPs por cromossomo e por gene e a lista de todos os SNPs que foram utilizados encontram-se na tabela suplementar 5.

4 Análises dos dados

4.1 Análises de SNPs em diferentes grupos populacionais

Inicialmente, verificamos a frequência alélica geral dos 408 SNPs na amostra de 333 indivíduos. Então foi testada associação entre os grupos das várias categorias populacionais previamente definidas e as frequências alélicas para todos os SNPs. Para isso, foram feitos testes exatos de Fisher com correção FDR BH (Benjamini e Hochberg, 1995).

As análises foram executadas com o *software* PLINK v. 1.9 (disponível em <https://www.cog-genomics.org/plink/1.9/>) (Purcell *et al.* 2007). Os gráficos foram plotados a partir do *software* Haploview v. 4.2 (disponível em <https://www.broadinstitute.org/haploview/downloads>) (Barrett *et al.* 2005).

Os testes de associação são uma investigação exploratória da relação entre SNPs dos genes de interesse e diferentes condições ambientais que poderiam interferir na seleção de fenótipos adaptativos. Por isso foram usados dados brutos, sem os filtros de controle de qualidade e de estrutura populacional comumente aplicados em estudos de associação caso-controle descritos em Marees (*et al.* 2018).

4.2 Análises de redes e combinações alélicas: Coeficiente de Correlação Personalizado e BlocBuster

O Coeficiente de Correlação Personalizado (CCC, do inglês, *Custom Correlation Coefficient*), é uma ferramenta matemática para identificar padrões de associação entre alelos de SNPs (Climer, *et al.*, 2014 a). O valor de CCC é calculado entre todos os pares de alelos de SNPs da amostra. Ele mede o desequilíbrio de ligação (LD, *linkage disequilibrium*) entre cada alelo de cada um dos SNPs, separando as combinações alélicas em elementos de vetor. Cada elemento do vetor indica o grau de associação entre alelos específicos. Às combinações alélicas são atribuídos pesos segundo a força com que se associam dentro de determinada amostra (Tiosano, *et al.*, 2016).

Como as associações alélicas podem abranger mais de um par de *loci*, Climer *et al.* (2014 a) acoplaram o CCC a outros programas em um *software* chamado BlocBuster (disponível em www.blocbuster.org). Ele extrai todos os elementos significativos de CCC e os agrega em redes. Desse modo, os alelos de SNP em LD na amostra, estejam eles no mesmo cromossomo ou não e independentemente do gene ao qual pertençam, serão mostrados interconectados nas redes multi-loci. Tais redes são compostas por nós ou vértices, alelos de SNP correlacionados, e arestas, valores de CCC significativos. O BlocBuster assume que os SNPs sejam bialélicos, portanto as redes são construídas com nós $2n$, onde n é o número de SNPs. A comparação entre cada dois SNPs bialélicos resulta em 4 valores, os que forem significativos são conectados pelas arestas (Climer, *et al.*, 2014 a e b; Tiosano, *et al.*, 2016).

O objetivo ao aplicar este método foi identificar combinações de alelos em desequilíbrio de ligação em diferentes genes que contribuem para um determinado fenótipo. Neste caso especificamente, buscamos compreender o papel dos genes candidatos selecionados quanto a sua contribuição para a adaptação das populações nativas americanas aos seus respectivos ambientes.

4.2.1 Filtro de desequilíbrio de ligação (LD)

Para chegarmos a uma lista final de SNPs a serem analisados juntos, tivemos que introduzir algumas estratégias. Em especial um filtro para detectar desequilíbrio de ligação (LD), como forma de retirar o excesso de variantes em *linkage*, herdadas juntas devido à proximidade física. Cada conjunto de SNPs em *linkage* representa, em termos estatísticos, aproximadamente a mesma informação, portanto uma única variante pode ser usada para representar cada região. Como o objetivo da investigação é buscar interações epistáticas entre alelos que interagem entre si para influenciar um fenótipo adaptativo, combinações redundantes não são informativas, podendo representar um ruído na análise.

Para gerar o subconjunto de SNPs aproximadamente em equilíbrio entre si, realizamos uma filtragem baseada no Fator de Inflação da Variância (VIF, do inglês *Variance Inflation Factor*). O VIF representa o incremento da variância devido à presença de correlação entre variáveis independentes. O procedimento foi feito a partir do *software* PLINK 1.07 (Purcell *et al.* 2007). Os parâmetros utilizados foram tamanho de janela de 50 SNPs, 5 SNPs para mudar a janela em cada etapa e limiar de 2 do VIF. Em cada etapa as variantes na janela que tiverem um VIF que exceda o limiar são removidas. Mais detalhes sobre o procedimento, consultar a documentação do PLINK 1.07 (<http://zzz.bwh.harvard.edu/plink/summary.shtml#prune>). A partir da aplicação deste filtro, foram excluídos 302 dos 408 SNPs, restando 106 SNPs.

4.2.2 Formação das redes

Com uso do programa BlocBuster, foi feito o cálculo de correlação personalizada, CCC, entre todos os alelos dos SNPs de todos os indivíduos da amostra. Portanto foi calculada a correlação entre os de 212 alelos dos 17 genes para os 333 nativos americanos. Os valores acima de 0,65 foram considerados significativos. O valor de 0,65 foi escolhido porque as correlações intergênicas começaram a aparecer a partir deste limiar. Este mesmo valor de significância para o CCC foi adotado por Tiosano (*et al.*, 2016) e permitiu a identificação de correlações intergênicas entre os SNPs genotipados para

aquele estudo. O procedimento de validação do limiar será explicado no próximo tópico (4.2.3).

A seguir aplicamos o BFS (*Breadth-First Search*), programa do BlocBuster que concatena em *clusters* os vértices que estão unidos por alguma aresta adjacente. Esses *clusters* ou blocos multi-alelo, são as redes. O arquivo de saída do CCC foi adaptado por nós para gerar os gráficos das redes no *software* R a partir do pacote Igraph.

4.2.3 Testes de permutação e validação do limiar

Para interpretar a validade do limiar de 0,65, fizemos 10 mil testes não paramétricos de permutação. Esses testes têm o objetivo de simular a hipótese nula de nenhuma associação, portanto são permutados aleatoriamente os genótipos de todos os indivíduos da amostra. Essa abordagem mantém a frequência original dos alelos da amostra. Com isso são avaliadas as correlações inerentes ao conjunto de dados para eliminar erros do tipo I. Este procedimento informa os valores limite para associação espúria. Com a amostra permutada, todos os valores de CCC acima do limiar foram considerados falso-positivos.

O número de falso-positivos é sensível ao tamanho da amostra, que quanto maior for, menor será, devido ao aumento do poder estatístico. Em contraste, o número de falso-positivos tende a aumentar com aumento do número de SNPs da amostra. Portanto, esses valores variam a cada conjunto de dados. Esses testes resultam em uma lista com o número de falso-positivos surgidos em cada iteração. Tal lista informa como se distribuem cada uma das quantidades de arestas falso-positivas. A implementação dos testes de permutação foi feita com o uso de um *script bash*, desenvolvido para o presente trabalho.

4.2.4 Simulação e medida de associação das arestas

Os valores de falso-positivo indicam a probabilidade global de alguma aresta ser falsa. Eles não são capazes de informar acerca da relação entre cada um dos alelos das redes. Então foram realizadas 10.000 simulações com objetivo de medir a probabilidade de associação específica de cada dois alelos nas redes.

As simulações sorteiam primeiramente o número de arestas que será excluído no teste de acordo com a distribuição de falso-positivos das permutações. Em seguida ocorre o sorteio da aresta, ou das arestas que serão eliminadas. Este segundo sorteio é ponderado pela probabilidade específica de cada aresta ser excluída. Cada uma delas tem um valor de conexão proporcional à força de interação entre os dois alelos que ela conecta, portanto, a chance de ser sorteada é o inverso da sua força. Ao fim das 10.000 iterações é quantificada a frequência em que cada dois vértices permanecem conectados. Esse resultado representa o grau de confiança que se tem na associação entre diferentes alelos das redes. Para realizar este procedimento foi desenvolvido um *script* de R.

4.2.5 Comparação entre redes e combinações alélicas específicas e grupos populacionais

As populações foram agregadas nas categorias descritas no tópico 3.2. Foi determinada a frequência e contabilizado o número de indivíduos das categorias que apresenta cada uma das redes geradas pelo BlocBuster. Adicionalmente foram identificadas as frequências de indivíduos que apresentam combinações específicas de alelos em cada um dos grupos. Para avaliar o significado biológico da distribuição das redes e das combinações alélicas nos grupos de interesse, foram aplicados testes exatos de Fisher com correção FDR BH. Todas as análises foram feitas com o *software* R, por meio de um *script* desenvolvido para a presente dissertação.

4.3 Estimativa do nível de estruturação das categorias

O índice F_{st} foi usado como parâmetro para estudar a diferenciação genética entre as populações que compreendem cada uma das categorias. O F_{st} mede os efeitos da estrutura populacional dentro das categorias, estimando o nível de divergência genética entre as populações que as compõem. Esta medida permite avaliar o grau de homogeneidade das categorias em relação aos genes de interesse deste estudo. Pode também ser referido como o componente da variância que é devido as diferenças entre populações que compõe um determinado grupo. O índice F_{st} foi estimado com o método de Weir e Cockerham (1984), que é uma forma de medir, o chamado por Wright (1951), índice de fixação, que leva em conta o tamanho amostral. O F_{st} geral e de cada uma das categorias, foi estimado por meio do *software* PLINK v. 1.9 (disponível em <https://www.cog-genomics.org/plink/1.9/>) (Purcell *et al.*, 2007).

Resultados

Os resultados abaixo mostram comparações entre categorias, considerando os alelos das duas vias (pigmentação e metabolismo/controla da expressão da vitamina D).

5.1 Influência das práticas de subsistência e hábitos de dieta:

Foi aplicado o teste exato de Fisher com correção FDR BH para comparar as frequências dos 408 SNPs entre os grupos caçadores-coletores (119 indivíduos) e agriculturistas (214 indivíduos). A comparação das frequências alélicas, resultou em 13 SNPs diferindo significativamente entre grupos (valor $p \leq 0.0001$), conforme figura 5. As planilhas 2, 3 e 4 da tabela suplementar 6 mostram respectivamente os resultados do teste exato de Fisher, os valores p ajustados por sete métodos diferentes de correção para múltiplos testes e o equilíbrio de Hardy-Weinberg entre agriculturistas e caçadores-coletores.

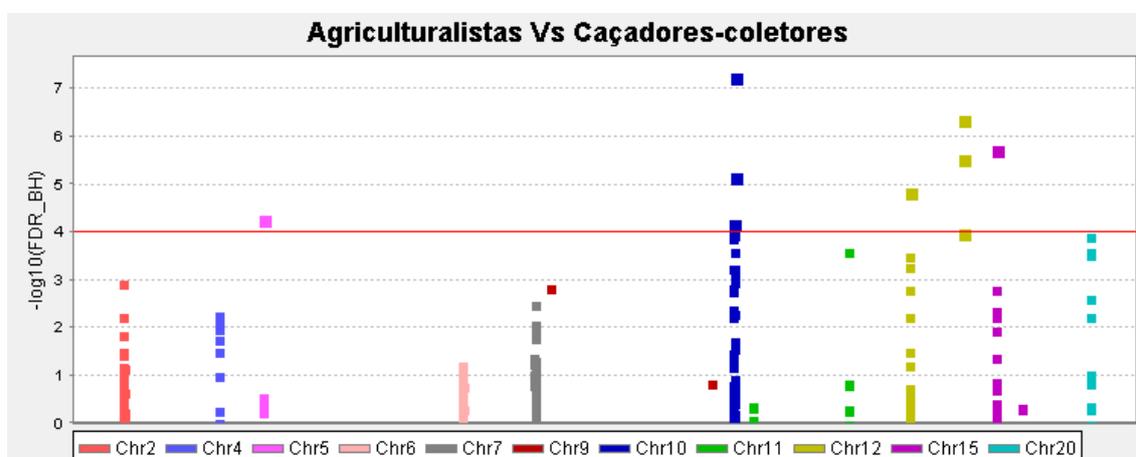


Figura 5- SNPs significantes encontrados na comparação entre agriculturistas e caçadores-coletores. As comparações foram feitas usando o teste exato de Fisher com correção FDR BH. Nós encontramos 13 SNPs significativos em 5 genes diferentes. Um SNP no gene *SLC45A2* (rs3776551, cromossomo 5). Quatro no gene envolvido no metabolismo de vitamina D, *CUBN* (rs4748341, rs3901939, rs3740164, rs10508517, cromossomo 10). Um no gene envolvido no controle da expressão da vitamina D, *VDR* (rs4516035, cromossomo 12). Seis SNPs no gene da via da pigmentação, *KITLG* (rs11104911, rs10777125, rs10506956, rs1703081, rs10777131, rs3782176, cromossomo 12). Um no gene da via da pigmentação, *OCA2* (rs4312267, cromossomo 15).

5.2 Influência da altitude:

Foi aplicado o teste exato de Fisher com correção FDR BH para comparar as frequências dos 408 SNPs entre os grupos terras altas (63 indivíduos) e terras baixas (109 indivíduos). A comparação das frequências alélicas, resultou em 5 SNPs diferindo significativamente entre grupos (valor $p \leq 0.0001$), conforme figura 6. As planilhas 5, 6 e 7 da tabela suplementar 6 mostram respectivamente os resultados do teste exato de Fisher, os valores p ajustados por sete métodos diferentes de correção para múltiplos testes e o equilíbrio de Hardy-Weinberg entre terras altas e terras baixas.

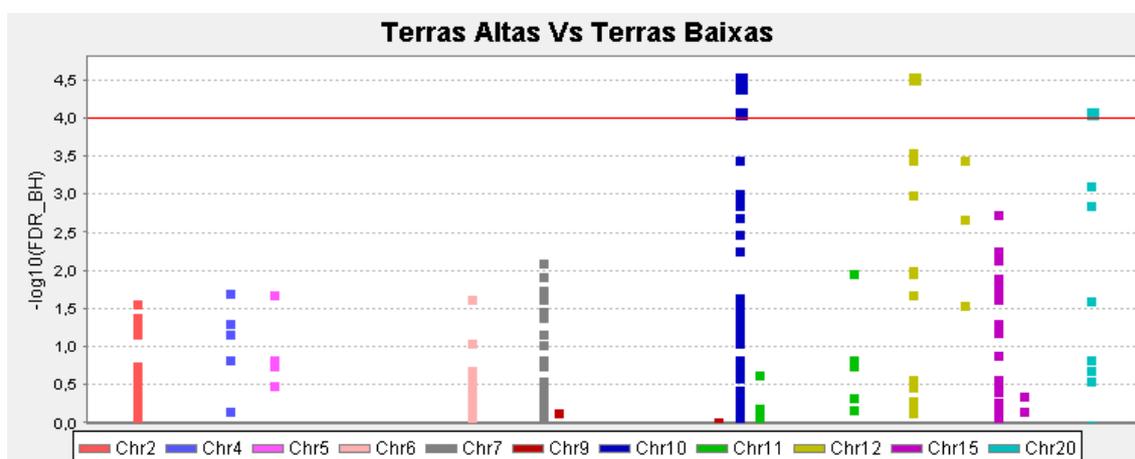


Figura 6- SNPs significantes encontrados na comparação entre terras altas e terras baixas. As comparações foram feitas usando o teste exato de Fisher com correção FDR BH. Foram encontrados 5 SNPs significantes em três genes diferentes. Três SNPs no gene envolvido no metabolismo de vitamina D, *CUBN* (rs3901939, rs4748341, rs7897550, cromossomo 10). Um no gene envolvido no controle da expressão da vitamina D, *VDR* (rs2239182, cromossomo 12). Um no gene envolvido no metabolismo de vitamina D, *CYP24A1* (rs3787557, cromossomo 20).

5.3 Distinção entre a influência da altitude e da dieta

Foi aplicado o teste exato de Fisher com correção FDR BH para comparar as frequências dos 408 SNPs entre os grupos agriculturalistas de terras baixas (151 indivíduos) caçadores-coletores (119 indivíduos). A comparação das frequências alélicas, resultou em 11 SNPs diferindo significativamente entre grupos (valor $p \leq 0.0001$), conforme figura 7. As

planilhas 8, 9 e 10 da tabela suplementar 6 mostram respectivamente os resultados do teste exato de Fisher, os valores p ajustados por sete métodos diferentes de correção para múltiplos testes e o equilíbrio de Hardy-Weinberg entre agriculturistas de terras baixas e caçadores-coletores.

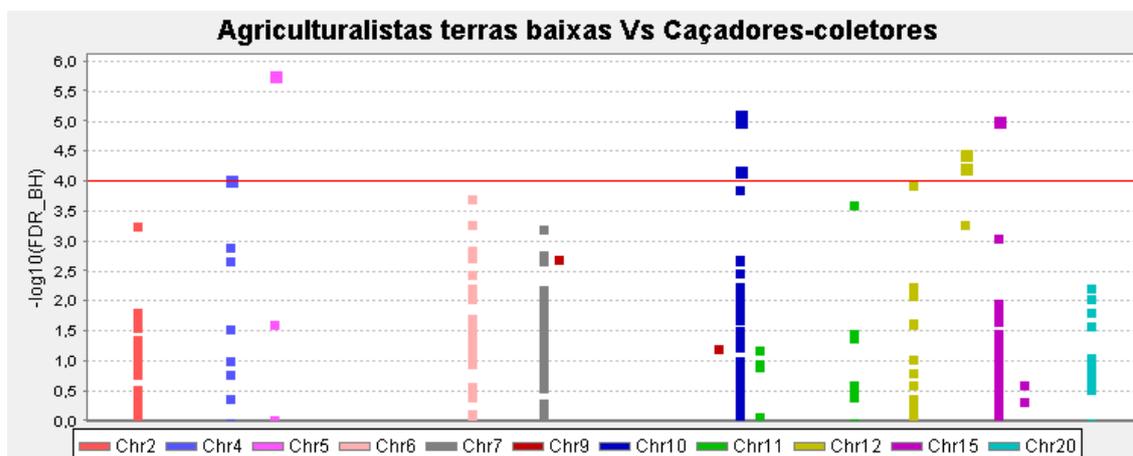


Figura 7- SNPs significantes encontrados na comparação entre agriculturistas de terras baixas e caçadores-coletores. As comparações foram feitas usando o teste exato de Fisher com correção FDR BH. Nós encontramos 11 SNPs significativos em 5 genes diferentes. Um SNP no gene envolvido no metabolismo de vitamina D, *GC* (rs4752, cromossomo 4). Um SNP no gene na via da pigmentação, *SLC45A2* (rs3776551, cromossomo 5). Três no gene envolvido no metabolismo de vitamina D, *CUBN* (rs4748341, rs3901939, rs11254325, cromossomo 10). Cinco no gene da via da pigmentação, *KITLG* (rs11104911, rs10777125, rs10506956, rs1703081, rs10777131 cromossomo 12). Um no da via da pigmentação, *OCA2* (rs4312267, cromossomo 15).

5.1.4 Influência da latitude:

Foi aplicado o teste exato de Fisher com correção FDR BH para comparar as frequências dos 408 SNPs entre os grupos Altas latitudes (14 indivíduos) e baixas latitudes (314 indivíduos). A comparação das frequências alélicas, resultou em 8 SNPs diferindo significativamente entre grupos (valor $p \leq 0.0001$), conforme figura 8. As planilhas 11, 12 e 13 da tabela suplementar 6 mostram respectivamente os resultados do teste exato de Fisher, os valores p ajustados por sete métodos diferentes de correção para múltiplos testes e o equilíbrio de Hardy-Weinberg entre altas latitudes e baixas latitudes.

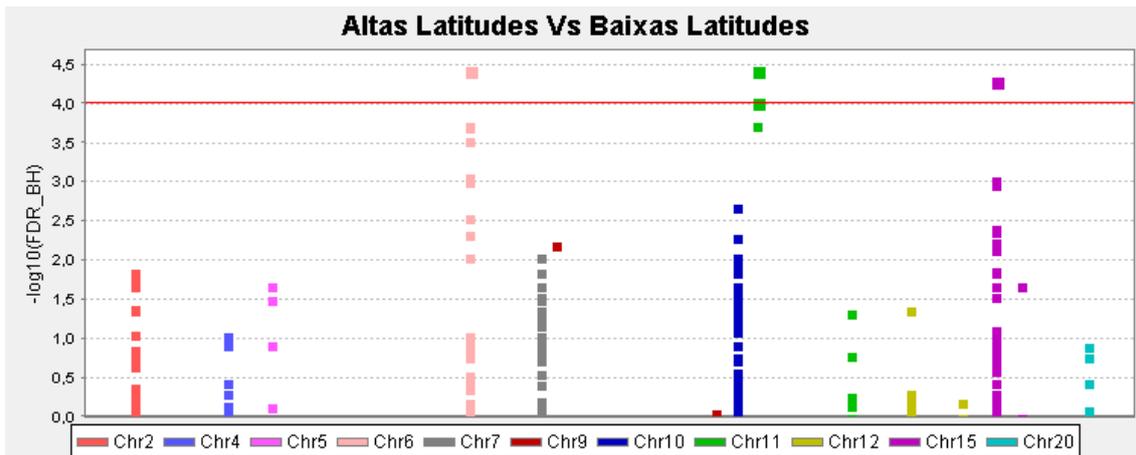


Figura 8- SNPs significantes encontrados na comparação entre altas latitudes e baixas latitudes. As comparações foram feitas usando o teste exato de Fisher com correção FDR BH. Foram encontrados 8 SNPs significativos em três genes diferentes. Três no gene da via da pigmentação, *OPRM1* (rs7759388, rs1918760, rs12212773, cromossomo 6). Quatro no gene envolvido no metabolismo da vitamina D, *CYP2R1* (rs1993116, rs2060793, rs1562902, rs16930625, cromossomo 11). Um no gene da via da pigmentação, *OCA2* (rs4778196, cromossomo 15).

6. Formação das redes e medida de associação entre as interações

6.1 Filtro de LD e aplicação do BlocBuster

A partir da aplicação do filtro de desequilíbrio de ligação que adotamos para o presente estudo, foram excluídos 302 dos 408 SNPs, restando 106 SNPs para as análises das redes. A aplicação do cálculo de CCC sobre os genótipos dos 212 alelos nos 333 nativos americanos, resultou em 34 valores acima do limiar de 0,65. Estes valores são representados por arestas que conectam alelos de SNPs, que constituem os vértices ou nós. Dos 212 alelos avaliados, 172 estão desconectados e 40 fazem parte das 9 redes geradas (Tabela Suplementar 7). A figura abaixo mostra as redes encontradas considerando nossa amostra como um todo.

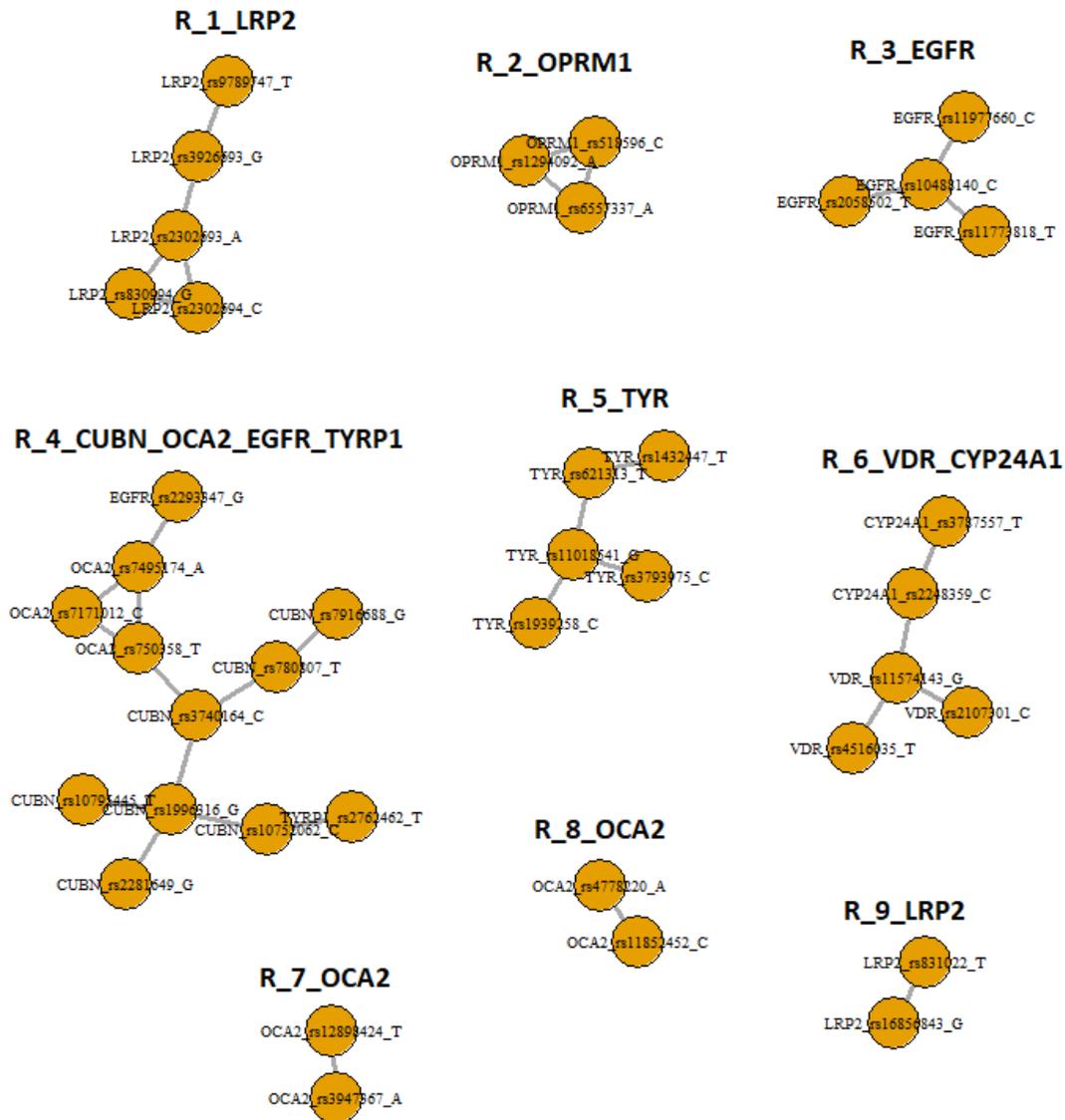


Figura 8- Representação das 9 redes alélicas identificadas pelo BlocBuster. Os círculos são os vértices que representam os alelos. Apenas os vértices conectados a arestas estão representados. As 34 arestas que conectam os vértices representam valores do coeficiente de correlação personalizado (CCC) $\geq 0,65$. Estas 34 arestas definem nove redes alélicas discretas numeradas de 1 a 9 conforme especificado na figura.

Sete das redes formadas não apresentam conexões intergênicas, ligando entre si apenas SNPs dos mesmos genes. Elas representam alelos ligados por proximidade física que restaram depois do filtro de desequilíbrio de ligação. As duas restantes são de particular interesse, *R_4_CUBN_OCA2_EGFR_TYRP1* e a *R_6_VDR_CYP24A1*, pois elas

conectam alelos de diferentes genes. A rede 4, mostra conexão entre um gene ligado ao metabolismo da vitamina D (*CUBN*), com três genes relacionados com a pigmentação (*OCA2*, *EGFR* e *TIRP1*). Em outras palavras, os alelos de *CUBN*, rs10752062/C, rs3740164/C, rs780807/T, rs1996316/G, rs10795445/T, rs2281649/G e rs7916688/G; de *OCA2* rs7171012/C, rs7495174/A e rs750358/T; de *EGFR* rs2293347/G; e de *TYRP1* rs2762462/T parecem estar sendo segregados juntos na amostra. Da mesma forma que acontece na R_6, em que os alelos de *VDR* rs2107301/C, rs11574143/G, rs4516035/T; e de *CYP24A1*, rs2248359/C e rs3787557/T, estão agregados.

É possível que estas duas redes tenham implicância funcional, visto que suas interações podem representar epistasia entre alelos de diferentes genes e vias, o que será avaliado nos itens a seguir. Vale destacar aqui que o termo epistasia está se referindo a possível existência de interação entre alelos, cujo produto tem valor adaptativo agregado.

6.2 Testes de permutação e associação entre alelos considerando as 9 redes

Os resultados dos testes de permutação revelaram uma taxa de 0,9673 (9.673 em 10.000 testes) valores de CCC $\geq 0,65$ por teste. Isso significa, em média, quase uma aresta falso-positiva por teste. Como as redes são formadas por um conjunto de arestas, um falso-positivo não invalida as redes como um todo. Considerando as 34 conexões das 9 redes geradas com os dados originais dos nativos americanos, cada uma delas têm a chance geral de 2,8% ($0.9673/34=0.02845$) de ser um falso-positivo. A distribuição de arestas falso-positivas geradas por teste, encontra-se na tabela 2.

Tabela 2: Distribuição dos falso-positivos gerados pelos 10 mil testes de permutação entre os genótipos dos 333 indivíduos da amostra. O limiar usado para formação das arestas foi $CCC \geq 0.65$.

Número de falso-positivos	Número de testes de permutação com cada quantidade de arestas falso-positivas.	
0	3803	38,03%
1	3711	37,11%
2	1722	17,22%
3	587	5,87%
4	142	1,42%
5	27	0,27%
6	5	0,05%
7	0	0,00%
8	3	0,03%

6.3 Interações intergênicas (alelo-alelo):

As conexões entre alelos ligados diretamente tendem a ser mais fortes do que aquelas entre alelos ligados indiretamente, por intermédio de outros alelos. Tendo isso em mente, passamos a avaliar a confiança estatística entre as conexões intergênicas diretas (alelo-alelo), com possível significado biológico. Para isso, foram feitas simulações de eliminação de arestas. Essas simulações tendem a eliminar conexões mais fracas, conforme descrito no item “4.2.4 Simulação e medida de associação das interações”, presente na sessão Amostra e Métodos.

Abaixo está a tabela que mostra o percentual de simulações em que as interações com possível interesse biológico se mantiveram nos testes de permutação.

Tabela 3: Percentual de simulações em que cada uma das interações intergênicas se manteve em 10.000 simulações.

Interação	Mantida em (%) simulações	Gene/SNP/Alelo	Gene/SNP/Alelo	Valor de CCC
<i>VDR_CYP24A1</i>	97,06%	<i>VDR</i> /rs11574143/G	<i>CYP24A1</i> /rs2248359/C	0,650575
<i>TYRP1_CUBN</i>	97,22%	<i>TYRP1</i> /rs2762462/T	<i>CUBN</i> /rs10752062/C	0,657256
<i>CUBN_OCA2</i>	96,89%	<i>CUBN</i> /rs3740164/C	<i>OCA2</i> /rs750358/T	0,651034
<i>OCA2_EGFR</i>	96,79%	<i>OCA2</i> /rs7495174/A	<i>EGFR</i> /rs2293347/G	0,653071

As simulações quantificam a probabilidade relativa de ligação entre cada uma das combinações alélicas de interesse. O número de simulações em que as conexões intergênicas se manteve revela uma grande confiança estatística nessas interações. Destaque para os alelos do gene *CUBN* (metabolismo de vitamina D) rs3740164/C e rs10752062/C que parecem estar segregando juntos com os alelos rs750358/T de *OCA2* e rs2762462/T de *TYRP1*, respectivamente, ambos genes ligados a via da pigmentação.

6.4 Comparação entre redes e combinações alélicas específicas e grupos populacionais

Foi aplicado um teste exato de Fisher com correção FDR BH para avaliar se as diferenças de frequência entre indivíduos das categorias previamente selecionadas eram estatisticamente significativas. Este procedimento foi feito tanto para as redes de interesse, quanto para as conexões intergênicas (alelo-alelo). Em ambas as análises o nível de significância foi fixado em $\alpha = 5\%$ ($p \leq 0.05$). Portanto, foi avaliado se há indício de coadaptação dirigida pelos diferentes contextos que potencialmente interferem na pressão seletiva do fenótipo de interesse. Informações mais detalhadas sobre cada um dos testes, encontram-se na tabela suplementar 8 (arquivo 'S.8_Redes_interações.xlsx').

6.4.1 Influência das práticas de subsistência e hábitos de dieta:

A comparação das redes de interesse entre as categorias agriculturistas e caçadores-coletores resultou em diferenças significativas em ambas. A rede inter-vias *R_4_CUBN_OCA2_EGFR_TYRP1* e a rede inter-genes *R_6_VDR_CYP24A1*, apresentaram os valores *p* respectivos de 0.025572 e 0.02202. Esse resultado indica que algumas combinações alélicas comuns no continente americano, em rede, diferem significativamente entre as categorias que discriminam as populações de acordo com práticas de subsistência e hábitos de dieta.

Já que as redes como um todo tem chances consideráveis de compreender algum falso-positivo, avaliaremos também as conexões intergênicas nas quais se tem bastante confiança estatística como já comentado acima (item 6.3).

Quanto às interações intergênicas diretas (alelo-alelo), as que diferem significativamente em frequência entre as categorias agriculturistas e caçadores-coletores são *CUBN_OCA2* (rs3740164/C_rs750358/T) e *CYP24A1_VDR* (rs2248359/C_rs11574143/G), cujos valores *p* são 0.044037 e 0.046044, respectivamente. Note que no primeiro caso são alelos de vias diferentes (metabolismo de vitamina D e via da pigmentação, respectivamente), e no segundo um gene do metabolismo da vitamina D e outro do controle da expressão da vitamina D.

6.4.2 Influência da altitude:

Na comparação entre as categorias terras altas e terras baixas, apenas a rede *R_6_VDR_CYP24A1* mostrou diferença estatisticamente significativa, com o valor *p* de 0.003336. A interação entre os alelos *CYP24A1_VDR* (rs2248359/C_rs11574143/G) desta rede foi a única que diferiu entre os grupos, com o valor *p* de 0.038300. *CYP24A1* é uma enzima catabólica, ela degrada a forma ativa da vitamina D₃. O *VDR*, por sua vez, é o receptor de vitamina D, principal mediador do controle da expressão gênica feito pelo hormônio.

6.4.3 Distinção entre a influência da altitude e das práticas de subsistência/hábitos de dieta:

A comparação entre as categorias agriculturalistas de terras baixas vs caçadore-coletores, considerando a quantidade de indivíduos com cada uma das redes intergênicas, resultou em diferença estatisticamente significativa na rede *R_4_CUBN_OCA2_EGFR_TYRP1*, com o valor $p = 0.004419$. As interações intergênicas (alelo-alelo), por sua vez, *CUBN_OCA2* (rs3740164/C_rs750358/T) e *CUBN_TYRP1* (rs10752062/C_rs2762462/T), ambos na rede inter-vias *R_4*, diferem entre grupos com os valores p respectivos de 0.034724 e 0.032064. Enquanto *CUBN* codifica o receptor renal que endocita o complexo da DBP-vitamina D₃, *OCA2* e *TYRP* atuam na rota da pigmentação.

6.4.4 Influência da latitude:

Ao comparar a ocorrência das redes entre indivíduos das categorias altas latitudes e baixas latitudes, não foi encontrada nenhuma diferença estatisticamente significativa com relação as redes. No entanto, duas das interações alélicas diferem entre os grupos, a *CUBN_OCA2* (rs3740164/C_rs750358/T; da rede *R_4*) com o valor p 0.000692 e a conexão também pertencente a rede *R_4*, *OCA2_EGFR* (rs7495174/A_rs2293347/G), valor p 0.034138.

7. Análises de SNPs em diferentes grupos populacionais

A frequência alélica geral dos 408 SNPs na amostra de 333 indivíduos, encontra-se na planilha 1 da tabela suplementar 6. O *Fst* obtido para cada uma das categorias está na tabela 4 (abaixo). Os valores mostram tendências gerais vistas em outros estudos, com diferentes marcadores genéticos. Isto é, grupos compostos de populações caçadoras-coletoras nativas americanas têm menor variabilidade genética dentro das populações e níveis altos de estruturação (medidos por índices como *Fst*), quando comparados aos números observados

em grupos autóctones de outros continentes (Bortolini *et al.*, 1997; Wang *et al.*, 2007). Isso indica baixo fluxo gênico, tamanho populacional efetivo pequeno, o que favorece a ação da deriva genética (mudanças devidas a flutuações casuais), pelo menos nos grupos nativos identificados com base na dieta e hábitos de vida, como caçadores-coletores. Por outro lado, nos grupos com os chamados agriculturalistas, sejam os de terras altas (andinos), quanto das baixas (mesoamericanos), o padrão é contrastante, com pouca estruturação (somente 2% do total da variância é devido as diferenças entre as populações que compõe a categoria, indicando tamanhos populacionais grandes e intenso fluxo gênico (Tarazona-Santos *et al.*, 2001). No caso dos Andes e Mesoamérica, esses achados são compatíveis com a história da região, onde ainda em tempos anteriores à chegada de Cristovão Colombo, impérios organizados e grandes cidades emergiram (Salzano e Bortolini, 2002). Esses achados demonstram que, como um todo, as distribuições alélicas nas categorias consideradas seguem o padrão esperado considerando marcadores genéticos neutros. Isso significa que eventos demográficos e casuísticos não podem ser descartados para explicar parte dos resultados apresentados. Não obstante, há achados instigantes, como o fato de a estruturação ser baixa dentro da categoria “agriculturalistas” que envolve andinos e mesoamericanos, o que sugere um fator comum levando a homogeneização nas distribuições alélicas. Esse fator comum poderia ser a agricultura.

Tabela 4: Valores de Fst considerando as categorias populacionais e os 408 SNPs localizados nos 17 genes candidatos.

Categoria	N° populações/ N° indivíduos	Fst (média)	Fst (estimativa ponderada)
Amostra total	33/333	0.0601629	0.0616636
Agriculturalistas	11/214	0.020581	0.020104
Caçadores-coletores	22/119	0.119347	0.124434
Terras altas	2/63	-0.00233535	-0.00206929
Terras baixas	20/109	0.11066	0.11845
Altas latitudes	3/14	0.0821329	0.110705
Baixas latitudes	28/314	0.0526703	0.0540097
Agriculturalistas das terras baixas	9/151	0.0222591	0.0232814

Discussão

8.1 Considerações sobre os métodos utilizados

A variabilidade de fenótipos complexos observados em populações humanas resulta de um conjunto de múltiplas variantes genéticas comuns e possivelmente algumas raras (Wainschtein *et al.*, 2019), bem como de suas conexões e possíveis efeitos epistáticos, podendo envolver várias vias (Gilbert-Diamond & Moore, 2011; Ritchie & Van Steen, 2018). Sendo assim, desvendar tais redes gênicas não é uma tarefa fácil. Considerando que muitos desses fenótipos têm valor adaptativo, o desafio torna-se maior ainda.

O termo epistasia foi introduzido por William Bateson (1909) para explicar as distorções na razão de segregação mendeliana, descrevendo o efeito de mascaramento de um gene sobre outro. Pouco depois, Ronald Fisher (1918) propôs um conceito estatístico de epistasia, definindo como os desvios da aditividade em um modelo estatístico linear. Este conceito permite medir o quanto os fenótipos observados diferem dos esperados devido a ação dos genes e suas variações (alelos) em conjunto.

A importância da epistasia estatística para a arquitetura genética de fenótipos complexos é inegável (Phillips, 2008; Gilbert-Diamond e Moore, 2011). Com base nessa ideia, nos últimos anos foram desenvolvidas uma enorme quantidade de ferramentas para detectar a interação gênica (Wei *et al.*, 2014; Ritchie e Van Steen, 2018). Os métodos mais tradicionais são baseados em regressão, como o Plink (com os comandos *epistasis* e *fast epistasis*) (Purcell *et al.*, 2007), e o BOOST (BOolean Operation-based Screening and Testing) (Wan *et al.*, 2010). Outra ferramenta comumente usada é o MDR (Multifactor Dimensionality Reduction) (Ritchie *et al.*, 2001), fundamentada na técnica de mineração de dados da inteligência artificial. Também existem os métodos que utilizam modelos Bayesianos, como o BEAM (Zhang e Liu, 2007), bem como os que se baseiam em desequilíbrio de ligação, como o BlocBuster (Climer, *et al.*, 2014b), utilizado neste trabalho.

Apesar da abundância de métodos que buscam epistasia, sua aplicação em estudos evolutivos é rara. Além disso, a epistasia, se extrapolada

para um contexto evolutivo se refere a possível existência de interação entre alelos cujo produto tem valor adaptativo agregado. Ou, conforme o conceito de “epistasia de *fitness*” (“*fitness epistasis*”), a interação entre alelos de diferentes *loci* em seus efeitos na aptidão. Esta definição implica em co-adaptação, termo cunhado por Dobzhansky (1948) para se referir a adaptação dependente de interações de aptidão entre alelos específicos em diferentes *loci*.

Ainda, as ferramentas citadas foram desenvolvidas com a demanda específica de entender a interação gênica no contexto de estudos caso-controle de doenças. Elas não consideram as peculiaridades do processo de adaptação de populações humanas. Portanto, podem requerer ajustes ou modificações para o uso nesses casos, além de muito cuidado na interpretação dos resultados obtidos. Por conta disso, vale à pena comentar sobre a forma como o método escolhido para realizar esta pesquisa foi aplicado em um artigo sobre adaptação no qual ele foi utilizado. Além disso, fizemos ajustes e adequações no método que também merecem ser comentados com maiores detalhes.

No artigo de Tiosano *et al.* (2016), os autores usaram o programa BlocBuster, do qual faz parte uma ferramenta matemática para detectar padrões de associação entre alelos de SNPs, denominada de Coeficiente de Correlação Personalizado (CCC; Climer, *et al.*, 2014a). Os autores fizeram uma triagem seletiva de variantes, genotipando 64 SNPs preferencialmente localizados em regiões sabidamente funcionais das vias de interesse e SNPs próximos a elas. No nosso trabalho foram utilizados dados de uma varredura genômica com informação sobre diferentes regiões de alguns dos genes candidatos, grande parte deles sem estudo prévio acerca das possíveis funções. Nessa varredura estão genotipados 408 SNPs dos genes de interesse. Para diminuir o número de variantes candidatas, aplicamos um filtro de desequilíbrio de ligação. Esse filtro retira o excesso de variantes em *linkage*, maximizando a chance de encontrar interações intergênicas sem perder muita informação.

Conforme mencionado, o BlocBuster é uma ferramenta que se baseia em um tipo de cálculo de desequilíbrio de ligação para detectar epistasia. Essa

classe de métodos parte do pressuposto de que SNPs que interagem para gerar um determinado fenótipo, uma doença por exemplo, devem coocorrer na amostra de indivíduos com esta doença em maior frequência do que na amostra sem ela (controle). Em outras palavras, a parcela dos casos da amostra seria enriquecida para tal interação entre SNPs em comparação aos indivíduos controle (Wei *et al.*, 2014). Um dos pontos centrais é estabelecer o que é considerado uma interação, ou seja, qual o limiar de coocorrência entre alelos de SNPs que permite considerar que eles interagem. Tiosano, *et al.* (2016) apresentam como critério de escolha do limiar a ausência de falso-positivos ou a presença de pouquíssimos deles em testes de permutação entre os genótipos da amostra. Os testes de permutação simulam a hipótese nula de nenhuma associação, então toda a associação que surge é espúria. Com o limiar estipulado em 0,65, os autores encontraram uma taxa de 0,137 falso-positivos por teste. Eles ressaltam a importância de manter uma baixa taxa de falso-positivos pelo fato de as ligações intergênicas serem, em geral, muito mais fracas do que as geradas por *linkage*. Portanto os falso-positivos provavelmente estariam na categoria das interações intergênicas, que são justamente as que importam para avaliar a epistasia entre diferentes rotas.

Após o desenvolvimento do *script* que nos permitiu realizar os testes de permutação, verificamos que a taxa de falso-positivos na nossa amostra, considerando o mesmo limiar, era bastante alta, 0,967. Isso produz uma grande chance de que redes com muitas arestas tenham algum falso-positivo. Entretanto, diferentemente de Tiosano e colaboradores (2016), o presente trabalho usou um filtro para retirar as ligações mais próximas. Sendo assim, todas as conexões têm aproximadamente a mesma força (tanto as intragênicas quanto as intergênicas). Por isso a ideia de não poder ter falso-positivos para creditar confiabilidade às conexões intergênicas não se sustenta no nosso estudo, conforme mostrado pelos testes de validação das interações, que levam em conta a força das arestas entre cada nó.

Além disso, Tiosano, *et al.* (2016) relataram que subindo o limiar para 0,7, no qual passa a não haver falso-positivos, também não há interações intergênicas. Portanto, o critério estabelecido pelos autores para escolha do limiar não tem o rigor de uma regra estrita. Outro ponto a ser considerado em

relação à alteração no critério de escolha e forma de validação do limiar é a sua variação nas diferentes aplicações do BlocBuster. Em cada trabalho que se aplica o cálculo CCC, é utilizado um critério diferente de escolha do limiar. No artigo que introduz o método (Climer *et al.*, 2014a), foram investigados 219.407 SNPs numa amostra de casos (N=74) e controles (N=70) para estudar a suscetibilidade à doença cardíaca hipertensiva. Foi feita uma comparação entre os valores de CCC dos indivíduos saudáveis (controles) e da amostra aleatorizada, na qual toda associação que surge é espúria. Os valores começaram a divergir a partir do limiar de 0,68. Então os autores adotaram o limiar de 0,7. Até 0,67 o número de conexões falsas e reais era aproximadamente o mesmo, depois disso as conexões passaram a surgir predominantemente na amostra real (não aleatória). Desse modo, os autores assumiram que conexões surgidas a partir do limiar de 0,7 eram reais. Em um trabalho sobre psoríase (Climer *et al.*, 2014b) foi escolhido o número de arestas igual ao número de SNPs. Depois foi observado que com a amostra aleatorizada não surgia nenhuma conexão tão forte quanto qualquer aresta da amostra não-permutada, portanto o limiar foi considerado válido. Considerando o que foi exposto acima, percebe-se que há uma certa flexibilidade no critério de escolha do limiar. Desse modo, pretendemos com nossa abordagem introduzir mais uma alternativa para a validação das interações.

Outra questão que cabe salientar em relação a implementação específica do método neste trabalho é a aplicação do cálculo de desequilíbrio de ligação sobre toda a amostra, ao invés de formar redes dentro das categorias. A adoção desta estratégia como abordagem inicial teve a intenção de replicar o protocolo estabelecido no trabalho de Tiosano, *et al* (2016). A medida conjunta de toda a amostra revela a tendência global das interações nos nativos americanos, em detrimento de conexões alélicas específicas de cada categoria. Por isso, interações fracas na amostra geral ou conexões específicas de categorias com poucos indivíduos, tendem a ser pouco representadas nesta análise.

É sabido que o desequilíbrio de ligação global medido em populações agrupadas pode gerar resultados que refletem a subdivisão populacional (Li e Nei, 1974). Qualquer força que induza diferença de frequência entre

subpopulações pode gerar LD na amostra agrupada. Buscamos especificamente detectar associações de alelos que reflitam possíveis processos de coadaptação. Entretanto, os padrões de LD das populações são fortemente afetados pela sua história demográfica particular (Laan e Pääbo, 1997). Como a coadaptação é apenas um dos fatores capazes de gerar associação entre alelos, é necessário ter muito cuidado ao interpretar os resultados.

Posteriormente, realizamos algumas tentativas de identificar redes dentro das categorias. Essa análise tinha o objetivo de otimizar a possibilidade de detecção de interações epistáticas específicas de cada categoria populacional. Estipulamos 10% como uma taxa aceitável de falso-positivos por teste, evitando a necessidade de validação posterior das interações. Com uso dos 106 SNPs, resultantes do filtro de LD aplicado sobre a amostra toda, as categorias testadas não mostraram conexões intergênicas. Para otimizar a possibilidade de detectar interações, testamos aplicar o filtro de LD dentro de uma das categorias e repetir o procedimento. Encontramos dificuldade no processo de validação de um limiar que seja informativo para o fim que buscamos (com conexões intergênicas) e ao mesmo tempo consistente estatisticamente. Por exemplo, aceitando cerca de 10% de falso-positivos por teste, apareceu apenas uma conexão intergênica na categoria testada. Por fim, concluímos que o estabelecimento de um protocolo viável demandaria mais tempo do que o disponível para terminar a dissertação. Sendo assim, optamos por deixar para estudos futuros abordagens dessa natureza.

8.2 Variações nos genes e vias sob investigação e suas implicações

8.2.1 Práticas de subsistência e hábitos de dieta

A comparação das frequências alélicas, considerando os SNPs candidatos, mostrou que há diferenças estatisticamente significativas entre estas, considerando aqueles localizados em alguns genes do

metabolismo/controla da expressão da vitamina D (*CUBN* e *VDR*), e em alguns dos genes da via da pigmentação de pele (*SLC45A2*, *KITLG* e *OCA2*).

Alguns casos merecem destaque. Por exemplo, SNPs (rs4748341 e rs3901939) do gene *CUBN* tem distribuições alélicas significativamente diferentes em duas análises (agriculturalistas das terras baixas vs caçadores-coletores; agriculturalistas vs caçadores-coletores), indicando que a diferença pode se dar devido às práticas de subsistência e hábitos de dieta. Conforme mencionado anteriormente, a cubilina, produto proteico deste gene, é um receptor das células do túbulo proximal dos rins. Ela exerce um papel importante na reabsorção da proteína de ligação à vitamina D, DBP. O processo de endocitose do complexo DBP-25(OH)D₃ está essencialmente envolvido na conversão de calcidiol a calcitriol, forma ativa do hormônio. Disfunções na cubilina foram associadas a perda urinária da DBP e diminuição dos níveis plasmáticos de 25(OH)D₃ e 1,25(OH)₂D₃ (Kaseda *et al.*, 2011). Portanto alterações nesse gene podem implicar em diferenças na mineralização óssea e outros processos nos quais o calcitriol está envolvido. Infelizmente não há estudos que testaram o potencial funcional das variantes intrônica que diferiram em frequência entre os grupos.

As frequências dos alelos do SNP (rs4312267) do gene *OCA2* diferem em duas análises (agriculturalistas das terras baixas vs caçadores-coletores; agriculturalistas vs caçadores-coletores). Parece que a alta altitude (de parte da amostra de agriculturalistas, já que ali estão os andinos) não está influenciando tais distribuições, reforçado ainda pelo fato de não aparecer diferença na comparação entre terras altas vs baixas. O mesmo poderia ser evocado para explicar o que acontece com os alelos de *KITLG* (rs11104911, rs107777125, rs10506956, rs1703081 e rs10777131), pelo menos considerando as categorias analisadas aqui. *KITLG* é uma proteína com amplo espectro de atuação, incluindo melanogênese. Um SNP nesse gene (rs1881227) já foi associado à pele mais clara e possivelmente estava presente no ancestral comum das populações asiáticas e europeias (Beleza, *et al.*, 2012; Deng e Xu, 2018; Stoneking, 2017). Outros estudos mostram que o alelo derivado é também bastante frequente nas populações ameríndias por eles amostradas (Deng e Xu, 2018). Embora não tenha entrado como candidato do nosso

estudo, por não constar na lista inicial de SNPs gerada pelo NCBI, o SNP rs1881227 de *KITLG* foi genotipado na nossa amostra. Nenhum dos 333 nativos americanos amostrados apresenta o alelo ancestral (C) deste SNP, indicando um cenário próximo da fixação do alelo derivado em populações autóctones da América. Além do SNP rs1881227, o SNP rs642742, localizado a 326 kb a montante do local de início da transcrição de *KITLG* apresenta a frequência do alelo ancestral >90% em africanos, similar à frequência do alelo derivado em europeus e asiáticos orientais, sugerindo que tal distribuição não pode ser facilmente explicada somente por eventos estocásticos e demográficos (Beleza, *et al.*, 2012). Sobre este SNP em particular não há informação na nossa varredura genômica. Também aqui, pelo menos até o momento, não existe qualquer estudo que indique funcionalidade nos *KITLG* SNPs que diferem nas categorias populacionais nativas americanas testadas na presente dissertação.

Por outro lado, dentre os alelos que diferem em suas frequências entre agriculturistas e caçadores-coletores, estão aqueles do SNP rs4516035. Este SNP está localizado no promotor do *VDR*, 1012 pb à montante do início do sítio de transcrição, éxon 1a, por isso é também referido como A-1012G (Halsall *et al.*, 2004). Esse SNP afeta a atividade transcricional do *VDR*, que é mais alta com a presença do alelo A (T) em comparação com o alelo G (C) (Fang *et al.*, 2005). É um polimorfismo muito estudado, variações nele foram associadas a diferenças na suscetibilidade à psoríase (Halsall *et al.*, 2005) e melanoma (Orlow *et al.*, 2011), por exemplo. Ele também exerce um papel no sistema imune (Halsall *et al.*, 2004; Laplana *et al.*, 2014) e foi relacionado a diferenças na altura (d'Alésio *et al.*, 2005). Outros sinônimos para esse polimorfismo são GATA, pelo seu papel na regulação do fator de transcrição GATA-3, e EcoRV, devido ao nome da enzima de restrição que o cliva.

Na nossa amostra o alelo G (C) é mais frequente em caçadores-coletores (0,3655) do que em agriculturistas (0,1846). Podemos sugerir então que um alelo de *VDR* (alelo A (T)), potencialmente mais ativo, pode ter sido favorecido em agriculturistas, para compensar a menor disponibilidade de vitamina D na dieta, de modo a atenuar possíveis deformidades esqueléticas e defeitos na dentição, que emergiram à medida que as populações humanas

deixaram suas práticas e hábitos de caçadores-coletores e passaram a cultivar alimentos (Brickley *et al.*, 2017).

Então, no período de expansão da agricultura, além da seleção de alelos que conferem pele mais clara (Richards, *et al.*, 2003; Marciniak e Perry, 2017; Stoneking, 2017; Quillen *et al.* 2019), a seleção natural pode ter favorecido também alelos ligados diretamente ao controle da expressão da vitamina D. Lucock *et al.* (2015) já haviam encontrado uma relação entre a frequência de alelos específicos de polimorfismos funcionais no gene *VDR* e a latitude (usada como medida indireta da variação da incidência de radiação) em populações africanas e eurasiáticas. Tiosano *et al.* (2016) também demonstraram que diferenças na latitude influenciaram a seleção conjunta de variantes funcionais do *VDR* em interação epistática com genes da via da pigmentação de pele. Porém, nenhum dos estudos considerou práticas de subsistência e hábitos de dieta como variável de interesse.

As variações na coloração da pele humana seguem um padrão “clássico” há bastante tempo conhecido ligado às latitudes, potencialmente, relacionado à necessidade de proteção contra a radiação ultravioleta intensa (UV) *versus* a necessidade de usar UV para a síntese da vitamina D. Não obstante essa regra geral, há exceções, como o caso dos nativos americanos, onde alguns povos apresentam pele mais escura (Inuits) e outros mais clara do que o esperado (povos que habitam os trópicos, sejam em terras altas quanto baixas). Sendo assim, outros elementos devem ser considerados. Aqui, pela primeira vez, introduzimos nos agrupamentos populacionais considerados práticas de subsistência e hábitos de dieta, além de altitude e latitude como medidas indiretas de incidência de radiação. A pressão seletiva sofrida pelas populações no processo de adaptação envolve tanto a produção endógena de vitamina D, quanto a sua disponibilidade na dieta. Nossos achados com o SNP funcional *VDR* rs4516035 ilustra bem a importância de considerar esse tipo de variável em estudos dessa natureza.

Considerando as interações, os alelos intergênicos em LD que diferiram de modo significativo em frequência foram *CUBN_OCA2* (rs3740164/C_rs750358/T), presente em 80% dos agriculturistas e 67% dos

caçadores-coletores e *CYP24A1_VDR* (rs2248359/C_rs11574143/G), que ocorre em 89% dos agriculturistas e 79% dos caçadores-coletores.

O primeiro caso envolve genes de vias diferentes, mas que parecem ter alelos sendo segregados juntos devidos a um possível efeito pleitrópico para um desfecho adaptativo que leve em conta um adequado metabolismo de vitamina D, cor da pele e práticas de subsistência. Não sabemos se os alelos *CUBN* rs3740164/C e *OCA2* rs750358/T são funcionais, tampouco o resultado de uma eventual epistasia entre eles. No entanto, é importante lembrar que uma dieta baseada em plantas domesticadas, normalmente tem baixas quantidades de vitamina D, em contraste com a alimentação de caçadores-coletores, mais rica e diversa, pois contém produtos de caça e recursos marinhos. As evidências arqueológicas indicam que a adoção de uma fonte de alimento menos variável advinda com a agricultura, sedentarização e urbanização não reduziram o estresse nutricional em comparação com caçadores-coletores em épocas pré-colombianas. Na verdade, o índice de saúde geral decaiu na medida em que as populações nativas americanas se tornaram sedentárias (Steckel e Rose, 2002). Esse período de transição dramático de práticas de subsistência e hábitos de dieta, possivelmente foi um momento importante para a ação da seleção natural e varreduras adaptativas, como mostra estudos envolvendo genes do metabolismo desenvolvido por nosso grupo de pesquisa (Hünemeier *et al.*, 2012).

Em relação ao segundo caso de interação gênica que difere em frequência entre as categorias aqui consideradas, *CYP24A1* rs2248359/C e *VDR* rs11574143/G, a comparação entre agriculturistas de terras baixas e caçadores coletores não resultou em diferença estatisticamente significativa. Caso a seleção tenha atuado para gerar essas diferenças de frequência, é provável que o fator altitude (diferença na incidência de radiação UV) tenha exercido papel predominante, pelo menos considerando esses genes e os *loci* em questão.

8.2.2 Influência da altitude

As análises para identificar as variantes que diferiam em suas distribuições nas categorias terras altas e terras baixas revelou alguns achados instigantes. Ao comparar a frequência de SNPs entre indivíduos das categorias aqui consideradas, foram encontradas diferenças significativas em alelos de três genes diferentes (*VDR*, *CYP24A1* e *CUBN*), todos na via da vitamina D.

O SNP rs2239182 localizado no gene *VDR*, é funcional, pois modula a expressão do fator de transcrição (Orlow *et al.*, 2015). Embora não tenha sido caracterizada precisamente a forma como as diferenças alélicas nesse polimorfismo atuam, ele foi associado a diversos fenótipos clínicos em diferentes populações, tais como suscetibilidade ao câncer de próstata (Oh *et al.*, 2014), diferenças no risco à fratura óssea (Yanovich *et al.*, 2012), miopia (Mutti *et al.*, 2011) e melanoma (Orlow *et al.*, 2015).

A única interação que diferiu significativamente entre grupos foi a *CYP24A1_VDR*, presente em 94% da amostra de terras altas e 79% dos indivíduos de terras baixas. Esse achado reforça o que foi dito no item anterior, que a interação *CYP24A1* rs2248359/C e *VDR* rs11574143/G parece ser influenciado por altitude, talvez incidência de UV.

O gene *CYP24A1* codifica a enzima 24-hidroxilase que degrada o metabólito ativo da vitamina D. O polimorfismo intrônico do *CYP24A1* (rs3787557), que difere entre terras altas e terras baixas, está localizado em uma região potencialmente funcional (Yao *et al.*, 2011). Ainda que sua função não tenha sido descrita, variações neste SNP foram relacionadas a doenças. Por exemplo, o alelo C em homozigose, foi associado a um aumento do risco a câncer de mama (Reimers *et al.*, 2014) e o alelo T, ao câncer de próstata (Holt *et al.*, 2014). Interessantemente, o alelo T deste SNP, aparece na rede *R_6_VDR_CYP24A1*, ligado ao alelo C do rs2248359 do mesmo gene. Este, por sua vez, participa da ligação intergênica com o alelo G do rs11574143 de *VDR*.

A ligação *CYP24A1_VDR* corrobora a ideia de que existe alguma diferença entre populações de terras altas e terras baixas no metabolismo e

controle da expressão da vitamina D. O rs2248359, está localizado na região promotora do *CYP24A1* e afeta a sua expressão, estando associado com diversas doenças relacionadas à deficiência de vitamina D. Estudos mostraram que a presença de pelo menos um alelo C neste locus é suficiente para regular positivamente a expressão da enzima *CYP24A1*, levando a diminuição dos níveis séricos de calcidiol (Pérez-Pérez, *et al.* 2018). Como salientado no parágrafo anterior, este alelo funcional de *CYP24A1* está ligado ao alelo G do rs11574143, localizado na região 3' UTR (3' não traduzida) de *VDR*. A função do SNP rs11574143 é desconhecida, contudo os alelos deste polimorfismo foram associados a doenças. Por exemplo, o alelo A deste SNP foi associado a risco aumentado a tuberculose pulmonar em uma população asiática (Hu *et al.*, 2016) e o G a câncer de próstata em homens com baixos níveis séricos de calcidiol, 25(OH)D₃ (Ahn *et al.*, 2009), o que corrobora a funcionalidade deste SNP. Polimorfismos no *VDR* podem influenciar na sua expressão e na estabilidade da proteína, visto que este fator de transcrição exerce múltiplas funções no organismo por isso qualquer alteração nele tem potencial impacto, tanto epistático quanto pleiotrópico.

Vale salientar, mais um vez, que as populações de terra altas da América do Sul (Andes) vivem em altitudes que podem chegar a mais de 4.000 metros, sendo expostas a níveis altos de radiação ultravioleta quando comparado ao nível do mar (para cada aumento de 1.000 metros na altitude, o nível de radiação UV aumentará cerca de 10-12%; https://www.who.int/uv/uv_and_health/en/). Vários estudos genéticos, incluindo de nosso grupo de pesquisa, tem revelado a existência de combinações alélicas específicas que dotaram essas populações a vida em altas altitudes, ambiente caracterizado por ar rarefeito (baixa quantidade de oxigênio disponível- hipóxia), baixa temperatura média anual, grande amplitude térmica, além da já comentada alta incidência de UV no ambiente.

Os Andinos têm a pele mais clara do que seria esperado para a incidência desse nível de UV no ambiente. Como exposto na revisão apresentada na introdução da presente dissertação, pele com pouca melanina tem menos proteção contra radiação ultravioleta, que degrada o folato,

elemento chave no reparo do DNA, protegendo contra defeitos severos como má formação do tubo neural. Por isso, ambientes com altas incidências de radiação costumam resultar em pressão seletiva para aumento da produção cutânea do pigmento melanina, que é fotoprotetivo, e consequente escurecimento da pele ao longo das gerações (Jablonski e Chaplin, 2000, 2010, 2017).

Embora saibamos que os Andinos têm a pele um pouco mais escura do que a maioria das demais populações autóctones da América, a discrepância em relação ao padrão geral para esse traço pode estar refletindo outras adaptações. Por exemplo, nosso grupo de pesquisa (Jacovas, *et al.*, 2015) mostrou a seleção de variantes de genes da via do *TP53*, que podem implicar em diferentes suscetibilidades ao câncer, inclusive o de pele. Também é possível que essas populações tenham desenvolvido alguma adaptação que confira mais eficiência às enzimas responsáveis pelo metabolismo do folato, hipótese não testada no presente trabalho. Outro possível resultado das altas exposições à radiação UV pode ser o relaxamento na pressão de seleção sobre variantes do metabolismo da vitamina D, resultando num metabolismo menos eficiente. Tendo em vista que peles com menos melanina produzem mais colecalciferol, uma variante mais eficiente do *CYP24A1*, que degrada a vitamina D, como visto aqui, não acarretaria nenhum risco de gerar qualquer problema a saúde destas populações de elevadas altitudes, visto a grande “oferta” de vitamina D nesse tipo de ambiente (com alta incidência de UV).

8.2.3 Influência da latitude

As categorias altas latitudes e baixas latitudes diferem quanto a frequência de SNPs de três genes. Um relacionado à síntese de vitamina D (*CYP2R1*) e os outros dois a pigmentação de pele (*OPRM1* e *OCA2*). As interações intergênicas que diferiram entre grupos foram *OCA2_EGFR* e *CUBN_OCA2*.

Populações de altas e baixas latitudes diferem em frequência em variantes do gene *CYP2R1*, que codifica a enzima responsável pela

hidroxilação do carbono 25 do colecalciferol (vitamina D), transformando-o em calcidiol, 25(OH)D₃. O SNP rs2060793 foi associado a diferenças nos níveis séricos de vitamina D em diferentes populações (Ahn *et al.*, 2010; Lu *et al.*, 2012; Zhang *et al.*, 2013; Sedky *et al.*, 2018). O alelo de risco para baixas concentrações de 25(OH)D₃ é o G (Zhang *et al.*, 2013), mais frequente em indivíduos de baixas latitudes (0,7245). Populações de nativos americanos que vivem em altas latitudes têm em alta frequência (0,75) o alelo A. Essa diferença faz sentido biológico, visto que em regiões de altas latitudes o alelo ligado a concentrações mais altas de calcidiol pode ter sido selecionado, dada a pouca disponibilidade de radiação (invernos longos, pouca luz solar durante o ano) para síntese do pró-hormônio.

Na avaliação de interações alélicas, a categoria altas latitudes têm em menor frequência (50%) a ligação OCA2_EGFR do que populações que habitam baixas latitudes (79%). Nesta conexão, estão o alelo A do rs7495174 do OCA2 e o alelo G do rs2293347 do EGFR. Ambos os genes são relacionados à cor, mas a função dos SNPs em questão não é conhecida. Um haplótipo contendo o alelo A (T) da variante do OCA2 foi fortemente relacionado a olhos claros e de modo mais sutil a cabelos e pele mais clara em indivíduos do norte da Europa. Entretanto, os SNPs daquele haplótipo muito provavelmente não são os responsáveis pela regulação do OCA2 que produz essa diferença de cor. Os autores do estudo acreditam que tais SNPs estejam em forte desequilíbrio de ligação com elementos regulatórios funcionais nas populações europeias em questão (Duffy *et al.*, 2006).

Vale lembrar que o fato de SNPs estarem em *linkage* em uma população, não implica que também estarão em outra. Isso ilustra o cuidado que se tem que ter ao interpretar variações associadas a um determinado fenótipo. No teste de associação entre SNPs, outra variante do OCA2, rs4778196, diferiu em frequência entre altas latitudes e baixas latitudes, este SNP não está ligado ao rs7495174 na amostra ($r^2= 0,0339$, $D'= 0,3977$) e, também, não há estudos que indiquem funcionalidade. Um dos alelos relacionado à cor mais escura de nativos americanos em relação aos europeus no estudo de Quillen, *et al.* (2012), o alelo A rs11238349 do gene EGFR,

encontra-se em altas frequências em toda a América, não tendo diferido entre categorias.

O outro gene com SNPs que diferem significativamente em frequência entre as duas categorias é *OPRM1*. O estudo que mostrou que diferenças em SNPs desse gene estavam relacionadas à cor mais escura em nativos americanos quando comparados aos europeus usou como amostra apenas populações americanas de altas altitudes (Quillen, *et al.*, 2012). Portanto é possível que se trate de uma adaptação específica das populações sujeitas a altos níveis de radiação UV. No entanto, a variante deste gene que foi associada a diferenças de cor entre europeus e nativos americanos no trabalho de Quillen e colaboradores (2012), rs6917661, não consta na nossa amostra. Também não há estudos que indiquem funcionalidade relacionada à cor nos SNPs que diferiram entre estas categorias (rs7759388, rs1918760 e rs12212773). Sendo assim, não é possível fazer inferências a esse respeito.

Outra questão a ser mencionada em relação à categoria “altas latitudes”, é que os Inuits apresentam diferenças importantes quando são comparados aos demais nativos americanos aqui amostrados. Há aproximadamente 16 mil anos, os nativos americanos separaram-se em dois ramos, nativos americanos do norte (NNA) e nativos americanos do sul (SNA). Diferentemente das demais populações consideradas, os Inuits são do grupo NNA (Moreno-Mayar *et al.*, 2018b). Além disso, os Inuits ilustram bem a importância do fluxo gênico constante no círculo polar Ártico, garantindo o contato entre povos da Ásia e América, mesmo depois do colapso da Beríndia (González-José *et al.*, 2007). Na presente dissertação, incluímos os Yaghan que habitam o extremo sul da América do Sul para maximizar a possibilidade de detectar algum sinal que não seja reflexo direto desta diferença de origem, mas sim de ambos viverem em climas extremos, próximos aos polos Norte e Sul, respectivamente. Desse modo, podemos especular que as diferenças de frequência entre as categorias denotem, portanto, convergência entre nativos americanos que vivem nesses ambientes extremos. Contudo, nossos achados relatados neste item devem ser vistos com bastante cautela, devido ao tamanho amostral pequeno das populações (Inuit e Yaghan) que representam a categoria “altas latitudes”.

Conclusões e reflexões finais

A relação entre os genes ligados à pigmentação de pele e à vitamina D é um tema de notável complexidade. A pigmentação de pele é caracterizada por grande heterogeneidade na arquitetura genética que a determina, o que reflete também uma das mais perceptíveis variações de fenótipos na nossa espécie. A diversidade de polimorfismos e genes envolvidos, suas relações epistáticas e o efeito pleiotrópico de algumas variantes que incidem no traço (cor da pele) ilustram as dificuldades e desafios enfrentados nesse tipo de estudo. A vitamina D, sua importância e implicações no desenvolvimento e na saúde humana, também é, por si só, um tema bastante complexo. Seu metabolismo inclui várias enzimas e o hormônio gerado incide em muitos processos biológicos, conforme pode ser visto pela quantidade de genes que o *VDR* regula.

Existe uma intrincada relação entre essas duas vias que varia entre populações. A exposição à radiação UV em pessoas com a pele mais escura tende a produzir menos vitamina D. Entretanto, populações de indivíduos com pele mais escura parecem usar o hormônio de modo mais eficiente do que aquelas com indivíduos de pele mais clara (Frost, 2009). Isso pode ocorrer de muitas formas diferentes. Dentre elas, maior absorção de cálcio e fósforo, melhor aproveitamento dessas moléculas, maior liberação de PTH, conversão mais eficiente às formas mais efetivas da vitamina D, transporte mais eficaz do metabólito através do sangue, receptor de vitamina D mais ativo, dentre outros (Wang *et al.*, 2010; McGrath *et al.*, 2010; Frost, 2009, 2012; Aslibekyan *et al.*, 2016).

Sobre o metabolismo da vitamina D em nativos americanos especificamente, não há muitas informações na literatura. A maior parte dos trabalhos tem como alvo de estudos os Inuits. Rejnmark *et al.* (2004) mostraram que, em comparação com dinamarqueses, os Inuits da Groenlândia têm níveis mais altos de calcitriol, porém níveis menores de calcidiol e do paratormônio (PTH), apesar de concentrações relativamente baixas de cálcio no plasma. O mesmo estudo sugere que os Inuits podem ter um “ponto de

ajuste” inerente menor para liberação de PTH regulada por cálcio ou uma produção aumentada de calcitriol renal.

Outro artigo documentou níveis séricos normais de cálcio e uma mais eficiente absorção desse elemento, apesar da baixa ingestão e de baixos níveis de calcidiol em crianças Inuits do Canadá. Os autores então postularam que a pressão seletiva teria favorecido receptores de vitamina D mais eficientes (genótipo *bb* do polimorfismo *BsmI* de *VDR*) (Sellers *et al.*, 2003). Em nossa amostra, considerando os 333 indivíduos de diferentes regiões da América, o alelo G (*b*) também tem alta frequência (0,92), indicando que uma eficiente absorção intestinal de cálcio é um padrão em toda a América.

Noutro estudo, com crianças nativas americanas que vivem em uma reserva a 43° de latitude, foi mostrado que 97% delas tinham baixas concentrações de vitamina D (<30 ng/mL) (Nsiah-Kumi *et al.*, 2012). Similarmente, outro trabalho, desta vez com indígenas argentinos de uma região de altas altitudes (aproximadamente 3.750 metros), mostrou haver índices deficitários ou insuficientes de vitamina D, apesar da grande exposição à radiação UV característica de elevadas altitudes (Hirschler *et al.*, 2013). Ou seja, é muito provável que baixas concentrações de vitamina D sejam uma característica geral de nativos americanos, não uma exclusividade de Inuits. É possível ainda que as diferenças fisiológicas dos Inuits, as quais viabilizam lidar eficientemente com as baixas concentrações de vitamina D, também não sejam uma idiosincrasia deles.

Interessantemente, esse dado vem na mesma direção de um estudo desenvolvido pelo nosso grupo de pesquisa com genes de metabolismo e populações nativas americanas, incluindo Inuits (Amorim *et al.*, 2017), que contrasta com os achados de Fumagalli *et al.* (2015). Os últimos autores, utilizando uma abordagem baseada em F_{ST} e dados de varredura genômica com uma amostra de Inuits da Groenlândia, sugeriram que nesse grupo populacional, ocorreu seleção positiva em genes codificadores das dessaturases de ácidos graxos (*FADS*), envolvidos no metabolismo de ácidos graxos poliinsaturados. Para eles, seria uma vantagem devido à dieta rica em gordura animal dos Inuits (Fumagalli *et al.*, 2015). Nosso estudo, por sua vez,

além dos Inuits considerou outros nativos americanos que vivem em uma gama diversa de ambientes. Demonstramos que a assinatura da seleção positiva em genes *FADS* estava presente em toda a América (Amorim *et al.*, 2017). Sugerimos então que esse evento de seleção estaria relacionado à adaptação local ao clima frio e à dieta rica em gordura animal que ocorreu na Beríngia, quando ali estiveram, por milhares de anos, os ancestrais de todos os nativos americanos.

Outro ponto a ser considerado, que certamente gerou implicações nos processos de adaptação às distintas regiões da América, é a cor de pele das populações que chegaram ao continente. Como já comentado, há sinal de seleção natural em alelos exclusivos de asiáticos nos genes da via da pigmentação *OCA2*, *DCT* (parálogo de *TYRP1*), e *MFSD12*, o que indica evolução paralela da despigmentação na Europa e Ásia (Deng e Xu, 2018; Adhikari *et al.*, 2019). Sabe-se que o alelo do gene *MFSD12* associado à cor clara em asiáticos também está presente em nativos americanos atuais, portanto chegou com eles à América (Adhikari *et al.*, 2019). A outra informação disponível até o momento, que nos permite inferir diretamente se as populações que chegaram tinham tom de pele equivalente ao atual, é a mutação nova do *OCA2* já mencionada. Nesse caso, o alelo derivado relacionado à cor clara não apresenta nenhum indício de seleção em nativos americanos (Donnelly, *et al.* 2012). Não obstante os poucos dados disponíveis, com base no que sabemos sobre pigmentação de pele em outras populações, é possível formular algumas hipóteses. Por exemplo, se a agricultura tiver de fato um papel relevante no clareamento de pele e o processo mais acentuado de despigmentação for relativamente recente, conforme o exemplo de populações europeias, é possível supor que as populações asiáticas que deram origem aos nativos americanos teriam a pele mais escura do que os asiáticos atuais.

Embora não possamos traçar mais do que hipóteses acerca dos dados acima expostos, essas são questões pertinentes para o entendimento geral tema. A forma como era o metabolismo da vitamina D ao entrar na América certamente impactou nos processos de adaptação a cada um dos ambientes com que se deparam os primeiros humanos que pisaram no continente

americano. Da mesma forma, a cor de pele dos indivíduos que deram origem a essas populações tem algum reflexo sobre a distribuição atual desse traço fenotípico no continente.

Ademais da complexidade geral do tema e das incertezas acima expostas, existem dificuldades de ordem metodológica relacionadas à busca de redes gênicas, já discutidas em mais detalhes anteriormente. Vale lembrar também que no presente estudo somente algumas variáveis ambientais (latitude e altitude) e de hábitos de subsistência/dieta (agriculturalista e caçador-coletor) foram consideradas para agrupar as populações nas distintas categorias, o que é obviamente uma simplificação. Ainda, utilizamos, por conveniência, um banco de dados genéticos já disponível para nativos americanos, o qual não contempla a maioria das populações autóctones ao longo do continente, nem tampouco muitos genes e variantes comuns nas duas vias de interesse.

A evolução é um processo inescapável, havendo tempo necessário, as populações com tamanho efetivo apropriado se adaptam, de um jeito ou outro. Não obstante as dificuldades e limitações, fomos capazes de detectar algumas diferenças entre regiões da América com distinta incidência de radiação UV. Por exemplo, maior atividade da 25-hidroxilase (gerada pelo alelo A do rs2060793), enzima responsável pela hidroxilação do carbono 25 do colecalciferol (vitamina D), em indivíduos de altas latitudes (e baixa incidência de UV), pode ter sido favorecida pela seleção natural. Por outro lado, a maior frequência da ligação entre alelos funcionais dos genes *CYP24A1* e *VDR* (alelo C do rs2248359 e alelo G do rs11574143) pode ter sido resultado do relaxamento da pressão de seleção para eficiência no metabolismo desse hormônio em ambientes com alta incidência de radiação UV em populações com a pele relativamente clara.

Com relação à alimentação, as diferenças de frequência no rs4516035 do *VDR* entre caçadores-coletores e agriculturalistas, podem indicar que um alelo mais ativo (alelo A) do gene que codifica o receptor de vitamina D teria sido favorecido em praticantes de agricultura, para compensar a menor disponibilidade dessa vitamina na dieta. A existência de diferenças de

frequência em polimorfismos de genes das vias da vitamina D e da pigmentação de pele entre populações em diferentes condições ambientais pode implicar atuação da seleção. Podemos pensar que há uma relação não casual entre essas diferenças de frequência e os ambientes das populações, especialmente no caso dos polimorfismos aqui citados, que são funcionais diferem conforme seria esperado para as populações das categorias estabelecidas.

Fomos ainda pioneiros em identificar potenciais alelos comuns (rs3740164/C e rs750358/T) de diferentes genes (*CUBN* e *OCA2*, respectivamente) pertencentes a vias distintas (metabolismo da vitamina D e rota da pigmentação) que podem estar sendo segregados em conjunto num contexto adaptativo que deve implicar um delicado balanço funcional, tanto epistático quanto pleiotrópico. Tal balanço deve contemplar um adequado metabolismo da vitamina D e do folato, concomitante com proteção contra formas de câncer de pele com relevância evolutiva para a nossa espécie por causarem morte ainda em idade reprodutiva, como o melanoma.

Esse tipo de cenário envolvendo redes de genes e de vias, embora esperado diante de fenótipos adaptativos complexos, não é fácil de ser revelado. Desse modo, acreditamos que o presente estudo possa ser o ponto inicial de outros, que trarão mais certezas e com isso mais conhecimento sobre os processos adaptativos que resultaram na bem-sucedida colonização humana na América, ainda em tempos pré-Colombianos.

Referências Bibliográficas

Adhikari K, Mendoza-Revilla J, Bortolini MC, Ruiz-Linares A, *et al.* (2019). A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nature Communications*. volume 10, Article number: 358 (2019)

Ahn, J., Albanes, D., Berndt, S. I., Peters, U., Chatterjee, N., Freedman, N. D., Prostate, Lung, Colorectal and Ovarian Trial Project Team (2009). Vitamin D-related genes, serum vitamin D concentrations and prostate cancer risk. *Carcinogenesis*, 30(5), 769–776. doi:10.1093/carcin/bgp055

Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K. C., McCullough, M. L., Gallicchio, L., ... Albanes, D. (2010). Genome-wide association study of circulating vitamin D levels. *Human molecular genetics*, 19(13), 2739–2745. doi:10.1093/hmg/ddq155

Alizadeh S, Djafarian K, Alizadeh H, Mohseni R, Shab-Bidar S. (2017) Common Variants of Vitamin D Receptor Gene Polymorphisms and Susceptibility to Coronary Artery Disease: A Systematic Review and Meta-Analysis. *J Nutrigenet Nutrigenomics* 2017;10:9-18. <https://doi.org/10.1159/000455914>

Amorim CEG, Daub JT, Salzano FM, Foll M and Excoffier L (2015) Detection of convergent genome-wide signals of adaptation to tropical forests in humans. *PLoS One* 10:e0121557.

Amorim, C.E.; Nunes, K.; Meyer, D.; Comas, D.; Bortolini, M.C.; Salzano, F.M.; Hunemeier, T. (2017) Genetic signature of natural selection in first Americans. *Proc. Natl. Acad. Sci. USA*, 114, 2195–2199.

Aoki K. (2002) Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Ann Hum Biol*, 29, pp. 589-608

Aslibekyan S, Vaughan LK, Wiener HW, *et al.* (2016) Linkage and association analysis of circulating vitamin D and parathyroid hormone identifies novel loci in Alaska Native Yup'ik people. *Genes. Nutr* 2016;11(23)

Barsh, G. S. (2003). What Controls Variation in Human Skin Color? *PLoS Biology*, 1(1), e27. <http://doi.org/10.1371/journal.pbio.0000027>

Bateson, W. (1909) *Mendel's Principles of Heredity* (Cambridge University Press, Cambridge, 1909).

Beleza, S., Santos, A. M., McEvoy, B., Alves, I., Martinho, C., Cameron, E., ... Rocha, J. (2012). The timing of pigmentation lightening in Europeans. *Molecular biology and evolution*, 30(1), 24–35. doi:10.1093/molbev/mss207

Benjamini Y e Hochberg Y. (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*, 57 (1995), pp. 289-300

Blum, H. F. (1961). Does the melanin pigment of human skin have adaptive value? *Quarterly Review of Biology*, 36, 50-63

Borradale e Kimlin, (2012). Folate degradation due to ultraviolet radiation: possible implications for human health and nutrition. *Nutr Rev.* 2012 Jul;70(7):414-22. doi: 10.1111/j.1753-4887.2012.00485.x. Epub 2012 May 22.

Bortolini MC, González-José R, Bonatto SL, Santos FR. (2014) Reconciling pre-Columbian settlement hypotheses requires integrative, multidisciplinary, and model-bound approaches *Proc Natl Acad Sci U S A.* 2014;111: E213–4. pmid:24398530

Bortolini MC, Salzano FM, Zago MA, Da Silva Júnior WA e Weimer T de A (1997) Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am J Phys Anthropol.*103(2):147-56.

Boxmeer, J. C., Smit, M., Utomo, E., Romijn, J. C., Eijkemans, M.J., Lindemans, J. et al. (2009). Low folate in seminal plasma is associated with increased sperm DNA damage. *Fertility & Sterility*, 92(2), 548–556

Brickley, M. B. D'Ortenzio, L. Kahlon B., Schattmann A. *et al.* (2017) Ancient vitamin D deficiency: longterm trends. *Current Anthropology* 58, no. 3 (June 2017): 420-427. <https://doi.org/10.1086/691683>

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

Cerqueira CC, Paixão-Côrtes VR, Zambra FM, Salzano FM, Hünemeier T, Bortolini MC. (2012) Predicting homo pigmentation phenotype through genomic data: from Neanderthal to James Watson. *Am J Hum Biol.*, 24: 705-709. 10.1002/ajhb.22263.

Chaplin G, Jablonski NG (1998) Hemispheric difference in human skin color. *Am J Phys Anthropol* 107:221–223, discussion 223–224

Chaplin G. (2004) Geographic distribution of environmental factors influencing human skin coloration. *Am J Phys Anthropol*, 125 pp. 292-302

Christakos S, Dhawan P, Verstuyf A, Verlinden L, Carmeliet G. (2016) Vitamin D: Metabolism, Molecular Mechanism of Action, and Pleiotropic Effects. *Physiol Rev.* 2016;96:365–408.

Cichorek, M., Wachulska, M., Stasiewicz, A., & Tymińska, A. (2013). Skin melanocytes: biology and development. *Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii*, 30(1), 30–41. <http://doi.org/10.5114/pdia.2013.33376>

Climer, S., Templeton, A. R., & Zhang, W. (2014 b). Allele-Specific Network Reveals Combinatorial Interaction That Transcends Small Effects in Psoriasis GWAS. *PLoS Computational Biology*, 10(9), e1003766. <http://doi.org/10.1371/journal.pcbi.1003766>

Climer, S., Yang, W., de las Fuentes, L., Dávila-Román, V. G., & Gu, C. C. (2014 a). A custom correlation coefficient (CCC) approach for fast identification of multi-SNP association patterns in genome-wide SNPs data. *Genetic Epidemiology*, 38(7), 610–621. <http://doi.org/10.1002/gepi.21833>

Crawford NG., Kelly DE., Hansen MEB., Beltrame MH., Fan S., Bowman SL., Jewett E., Ranciaro A., Thompson S., Lo Y., Pfeifer SP., Jensen JD., Campbell MC1., Beggs W., Hormozdiari F., *et al.* (2017) Loci associated with skin pigmentation identified in African populations. *Science* 358(6365).

D’Orazio, J., Jarrett, S., Amaro-Ortiz, A., & Scott, T. (2013). UV Radiation and the Skin. *International Journal of Molecular Sciences*, 14(6), 12222–12248. <http://doi.org/10.3390/ijms140612222>

d’Alésio A., Garabédian M., Sabatier J.P., Guaydier-Souquières G., Marcelli C., Lemaçon A., *et al.* (2005). Two single-nucleotide polymorphisms in the human vitamin D receptor promoter change protein-DNA complex formation and are associated with height and vitamin D status in adolescent girls. *Hum Mol Genet*, 14 (2005), pp. 3539-3548

Deng, L., & Xu, S. (2018). Adaptation of human skin color in various populations. *Hereditas*, 155, 1. <http://doi.org/10.1186/s41065-017-0036-2>

Diamond J. (1991). *The rise and fall of the third chimpanzee*. London, UK: Vintage Press. 573.2 ISBN 0-09-174268-4

Dobzhansky T. (1948). Genetics of natural populations; experiments on chromosomes of *Drosophila pseudoobscura* from different geographic regions. *Genetics*, 33(6), 588–602.

Donnelly MP, Paschou P, Grigorenko E, *et al.* (2012) A global view of the OCA2-HERC2 region and pigmentation, *Hum Genet.*, vol. 13.pg. 683-696

Duffy, D. L., Montgomery, G. W., Chen, W., Zhao, Z. Z., Le, L., James, M. R., ... Sturm, R. A. (2006). A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *American journal of human genetics*, 80(2), 241–252. doi:10.1086/510885

Elias, P. M., & Williams, M. L. (2013). Re-appraisal of current theories for the development and loss of epidermal pigmentation in hominins and modern humans. *Journal of Human Evolution*, 64(6), 687–692. <http://doi.org/10.1016/j.jhevol.2013.02.003>

Engelsen, O. (2010). The Relationship between Ultraviolet Radiation Exposure and Vitamin D Status. *Nutrients*, 2(5), 482–495. <http://doi.org/10.3390/nu2050482>

Fagundes NJ, Kanitz R, Eckert R, *et al.* (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet.* 2008;82(3):583-92.

Fang Y., van Meurs J.B., d’Alesio A., Jhamai M., Zhao H., Rivadeneira F., *et al.* (2005). Promoter and 3'-untranslated-region haplotypes in the vitamin

d receptor gene predispose to osteoporotic fracture: The Rotterdam study *Am J Hum Genet*, 77 (2005), pp. 807-823

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edin* 52:399–433

Frost P. Black-white differences in cancer risk and the vitamin D hypothesis. *J Natl Med Assoc.* 2009;101(12):13103.

Frost, Peter (2012) “Vitamin D Deficiency among Northern Native Peoples: A Real or Apparent Problem?”, *International Journal of Circumpolar Health*, 71: 18001. <https://www.tandfonline.com/doi/full/10.3402/ijch.v71i0.18643>

Fumagalli M, et al. 2015. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science.* 349(6254):1343–1347.

Gibbons A. (2007) European skin turned pale only recently, gene suggests. *Science.* 2007;316:364.

Gilbert-Diamond, D., & Moore, J. H. (2011). Analysis of gene-gene interactions. *Current protocols in human genetics*, Chapter 1, Unit1.14. doi:10.1002/0471142905.hg0114s70

González-José R., Bortolini M.C., Santos F.R., Bonatto S.L. (2008) The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am. J. Phys. Anthropol.*, 137 pp. 175-187

Greaves, M. (2014). Was skin cancer a selective force for black pigmentation in early hominin evolution? *Proceedings of the Royal Society B: Biological Sciences*, 281(1781), 20132955. <http://doi.org/10.1098/rspb.2013.2955>

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328: 710–22.

Guenther C.A. Tasic B. Luo L. Bedell M.A. Kingsley D.M. (2014) A molecular basis for classic blond hair color in Europeans *Nat. Genet* 2014 doi:10.1038/ng.2991

Halsall JA, Osborne JE, Pringle JH, Hutchinson PE. (2005) Vitamin D receptor gene polymorphisms, particularly the novel A-1012G promoter polymorphism, are associated with vitamin D3 responsiveness and non-familial susceptibility in psoriasis. *Pharmacogenet Genomics*, 15 (2005), pp. 349-355

Halsall, J. A., Osborne, J. E., Potter, L., Pringle, J. H., & Hutchinson, P. E. (2004). A novel polymorphism in the 1A promoter region of the vitamin D receptor is associated with altered susceptibility and prognosis in malignant melanoma. *British journal of cancer*, 91(4), 765–770. doi:10.1038/sj.bjc.6602006

Hancock, A. M., Witonsky, D. B., Alkorta-Aranburu, G., Beall, C. M., Gebremedhin, A., Sukernik, R., Di Rienzo, A. (2011). Adaptations to

climate-mediated selective pressures in humans. *PLoS Genetics*, 7, e1001375. <http://doi.org/10.1371/journal.pgen.1001375>

Haussler, MR, Jurutka, PW, Mizwicki, M, Norman, AW. (2011) Vitamin D receptor (VDR)-mediated actions of $1\alpha,25(\text{OH})_2$ vitamin D₃: genomic and non-genomic mechanisms. *Best Pract Res Clin Endocrinol Metab.*;25:543–559.

Hirschler V, Maccallini G, Molinari C, *et al.* (2013) Low vitamin D concentrations among indigenous Argentinean children living at high altitudes. *Pediatr Diabetes*. 2013;14(3):203–10.

Hochberg Z, Templeton AR. Evolutionary perspective in skin color, vitamin D and its receptor. *Hormones*. 2010;9:307–311.

Holick MF. (2016) Biologic effects of sunlight, ultraviolet radiation, visible light, infrared, and vitamin D for health. *Anticancer*; 36:1345-56

Holt, S. K., Kwon, E. M., Koopmeiners, J. S., Lin, D. W., Feng, Z., Ostrander, E. A., ... Stanford, J. L. (2010). Vitamin D pathway gene variants and prostate cancer prognosis. *The Prostate*, 70(13), 1448–1460. doi:10.1002/pros.21180

Hosseini-nezhad, A., & Holick, M. F. (2013). Vitamin D for Health: A Global Perspective. *Mayo Clinic Proceedings*. Mayo Clinic, 88(7), 720–755. <http://doi.org/10.1016/j.mayocp.2013.05.011>

Hu, Q., Chen, Z., Liang, G., Mo, F., Zhang, H., Xu, S., ... Jin, T. (2016). Vitamin D receptor gene associations with pulmonary tuberculosis in a Tibetan Chinese population. *BMC infectious diseases*, 16(1), 469. doi:10.1186/s12879-016-1699-4

Hünemeier T, Amorim CE, Azevedo S, Contini V, Acuña-Alonzo V, *et al.* (2012) Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. *PLoS One* 7: e38862.

Ito, S. and K. Wakamatsu (2003) Quantitative analysis of eumelanin and pheomelanin in humans, mice, and other animals: A comparative review. *Pigment Cell Res*. 16, 523–531.

Jablonski NG (2004) The evolution of human skin and skin color. *Annual Review of Anthropology* 33: 585–623.

Jablonski NG, Chaplin G (2000) The evolution of human skin coloration. *J Hum Evol* 39: 57–106.NG

Jablonski NG, Chaplin G (2010) Colloquium Paper: Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences of the United States of America* 107: 8962–8968.

Jablonski NG, Chaplin G. (2013) Epidermal pigmentation in the human lineage is an adaptation to ultraviolet radiation. *J Hum Evol.*;65:671-675. <http://dx.doi.org/10.1016/j.jhevol.2013.06.004>

Jablonski NG, Chaplin G. (2017) The colours of humanity: the evolution of pigmentation in the human lineage. *Philosophical Transactions of the Royal Society B: Biological Sciences.*;372:20160349. doi:10.1098/rstb.2016.0349

Jablonski NG., Chaplin G, (2009) Vitamin D and the evolution of human depigmentation. *Am J Phys Anthropol* 139: 451–461.

Jacovas VC, Couto-Silva CM, Nunes K, Bortolini MC, Hünemeier T, et al. (2018). Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. *Scientific Reports* volume 8, Article number: 12733

Jacovas VC, Rovaris DL, Pérez O, de Azevedo S, Macedo GS, Sandoval JR, Salazar-Granara A, Villena M, Dugoujon J-M, Bisso-Machado R et al. (2015) Genetic variations in the TP53 pathway in Native Americans strongly suggest adaptation to the high altitudes of the Andes. *PLoS One* 10:0137823.

Jolliffe DA, Walton RT, Griffiths CJ, et al. (2015) Single nucleotide polymorphisms in the vitamin D pathway associating with circulating concentrations of vitamin D metabolites and non-skeletal health outcomes: Review of genetic association studies. *J Steroid Biochem Mol Biol.*; pii: S0960-0760(15)30153-9. 10.1016/j.jsbmb.2015.12.007

Juzeniene, A., Setlow, R., Porojnicu, A. *et al.* (2009) Development of different human skin colors: a review highlighting photobiological and photobiophysical aspects *J Photochem Photobiol B*, 96, pp. 93-100

Kaseda R, Hosojima M, Sato H, Saito A 2011 Role of megalin and cubilin in the metabolism of vitamin D3. *Ther Apher Dial* 15(Suppl 1):14–17

Khouja, J., Lewis, S. J., & Bonilla, C. (2018). Influence of maternal and own genotype at tanning dependence-related SNPs on sun exposure in childhood. *BMC Medical Genetics*, 19, 62–69. <http://doi.org/10.1186/s12881-018-0575-z>

Kini U., Nandeesh B.N. (2012) Physiology of Bone Formation, Remodeling, and Metabolism. In: Fogelman I., Gnanasegaran G., van der Wall H. (eds) *Radionuclide and Hybrid Bone Imaging*. Springer, Berlin, Heidelberg

Kuhlwilm, M. *et al.* (2016) Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* 530, 429–433.

Laan, M., & Pääbo, S. (1997). Demographic history and linkage disequilibrium in human populations. *Nature Genetics*, 17(4), 435–438. doi:10.1038/ng1297-435

Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, Hughes D, Rohland N, Pilli E, Longo L, Condemi S, de la Rasilla M, Fortea J, Rosas A, Stoneking M, Schö"neberg T, Bertranpetit J, Hofreiter M. (2007) A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318:1453–1455.

Lewis B, Jurmain R, Kilgore L, (2013) Introduction to physical anthropology and Archaeology. Eleventh Edition. Wadsworth Cengage Learning.

Lowie R. H. American Culture History. (1940) Amer. Anthropol. 1940, 42: 409-428.

Lu L, Sheng H, Li H, Gan W, Liu C, Zhu J, Loos RJ, Lin X. (2012) Associations between common variants in GC and DHCR7/NADSYN1 and vitamin D concentration in Chinese Hans. Hum Genet. 2012; 131(3):505–12.

Lucock, M., Yates, Z., Martin, C., Choi, J.-H., Boyd, L., Tang, S., ... Veysey, M. (2014). Vitamin D, folate, and potential early lifecycle environmental origin of significant adult phenotypes. *Evolution, Medicine, and Public Health*, 2014(1), 69–91. <http://doi.org/10.1093/emph/eou013>

Lucock, M.; Jones, P.; Martin, C.; Beckett, E.; Yates, Z.; Furst, J.; Veysey, M. (2015) Vitamin D: Beyond metabolism. *J. Evid. Based Complement. Altern. Med.*, 20, 310–322.

Marciniak S., Perry GH. (2017) Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet.* 2017 Nov;18(11):659-674. doi: 10.1038/nrg.2017.65. Epub 2017 Sep 11.

Marees A. T., Kluiver H., Stringer S., Vorspan F, et al. (2018). A tutorial on conducting genomewide association studies: Quality control and statistical analysis, *International Journal of Methods in Psychiatric Research*, vol. 27, no. 2, Jun. 2018, [online] Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6001694/>.

Martin, AR., Lin M., Granka JM., Myrick JW., Liu X., Sockell A., Atkinson EG., Werely CJ., Iler MM., Sandhu MS., *et al.* (2017) An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell.* 2017 Nov 30;171(6):1340-1353.e14. doi: 10.1016/j.cell.2017.11.015.

Mathieson Iain, Lazaridis Iosif, Reich David, et al., (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503 (24 December 2015). doi:10.1038/nature16152

Mathur U, Datta SL, Mathur BB. (1977) The effect of aminopterin-induced folic acid deficiency on spermatogenesis. *Fertil Steril.* 1977;28:1356–1360.

McEvoy, B., Beleza, S. & Shriver, M.D. (2006) The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Hum. Mol. Genet.* 15, R176–R181

McGrath JJ, Saha S, Burne THJ, Eyles DW. (2010) A systematic review of the association between common single nucleotide polymorphisms and 25-hydroxyvitamin D concentrations. *J Steroid Biochem Mol Biol.* 2010;121(12):4717.

Moorjani P, Amorim CEG, Arndt PF, Przeworski M (2016) Variation in the molecular clock of primates. *Proc Natl Acad Sci USA* Sep 6 113 (38): 10607–10612.

Moreno-Mayar J. V., Potter B. A., Nielsen R., Meltzer D. J., Willerslev E. *et al.* (2018a). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* 553, 203–207 (2018). doi:10.1038/nature25173 Medline

Moreno-Mayar, L. Vinner, P. de Barros Damgaard, C. de la Fuente, J. Chan, J.P. Spence, M.E. Allentoft, T. Vimala, F. Racimo, T. Pinotti, *et al.* (2018b) Early human dispersals within the Americas. *Science* (2018), 10.1126/science.aav2621

Murray FG., Pigmentation, sunlight, and nutritional disease. (1934) *Am Anthropol*, 36 1934, pp. 438-445

Mutti, D. O., Cooper, M. E., Dragan, E., Jones-Jordan, L. A., Bailey, M. D., Marazita, M. L., ... CLEERE Study Group (2011). Vitamin D receptor (VDR) and group-specific component (GC, vitamin D-binding protein) polymorphisms in myopia. *Investigative ophthalmology & visual science*, 52(6), 3818–3824. doi:10.1167/iovs.10-6534

Nsiah-Kumi PA, Erickson JM, Beals JL *et al.* (2012) Vitamin D insufficiency is associated with diabetes risk in native American children. *Clin Pediatr (Phil)* 2012: 51: 146–153.

Nunes, S. B. R., de Matos Oliveira, F., Neves, A. F., Araujo, G. R., Marangoni, K., Goulart, L. R., & Araújo, T. G. (2016). Association of vitamin D receptor variants with clinical parameters in prostate cancer. *SpringerPlus*, 5, 364. <http://doi.org/10.1186/s40064-016-2009-8>

Oh JJ, Byun SS, Lee SE, Hong SK, Jeong CW, Kim D, et al. (2014) Genetic variations in VDR associated with prostate cancer risk and progression in a Korean population. *Gene*. 2014;533:86–93.

Orlow, I., Reiner, A. S., Thomas, N. E., Roy, P., Kanetsky, P. A., Luo, L., *et al.* GEM Study Group (2015). Vitamin D receptor polymorphisms and survival in patients with cutaneous melanoma: a population-based study. *Carcinogenesis*, 37(1), 30–38. doi:10.1093/carcin/bgv157

Orlow, I., Roy, P., Reiner, A. S., Yoo, S., Patel, H., Paine, S., ... GEM Study Group (2011). Vitamin D receptor polymorphisms in patients with cutaneous melanoma. *International journal of cancer*, 130(2), 405–418. doi:10.1002/ijc.26023

Park, S. Hongu, N. Daily J.W. (2016) Native American foods: History, culture, and influence on modern diets. *J Ethn Foods*, 3, pp. 171-177

Pérez-Pérez S, Domínguez-Mozo MI, GarcíaMartínez MÁ, Aladro Y, Martínez-Ginés M, García-Domínguez JM, et al. (2018) Study of the possible link of 25-hydroxyvitamin D with Epstein-Barr virus and human herpesvirus 6 in patients with multiple sclerosis. *Eur J Neurol* 2018;25:1446–1453

Phillips P. C. (2008). Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nature reviews. Genetics*, 9(11), 855–867. doi:10.1038/nrg2452

Pilz S, Verheyen N, Grübler MR, Tomaschitz A, März W. (2016) Vitamin D and cardiovascular disease prevention. *Nat Rev Cardiol*. 2016 May 6

Pinotti T., Santos F., Tyler-Smith, C. et al. (2019). Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and early Population structure of Native American Founder. *Cell*. Volume 29, Issue 1, 7 January 2019, Pages 149-157.e3

Posth C, Nakatsuka N, Lazaridis I, Skoglund P, Mallick S, Lamnidis TC, Rohland N, Nägele K, Adamski N, Bertolini E, et al. (2018). Reconstructing the deep population history of Central and South America. *Cell*. 2018;175(5):1185–97.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.S

Quillen E. E. (2015). The Evolution of Tanning Needs Its Day in the Sun. *Human Biology*, 87(4), 352. doi:10.13110/humanbiology.87.4.035

Quillen E. E., Norton H. L., Parra E. J. Jablonski N. G. (2019) Shades of complexity: New perspectives on the evolution and genetic architecture of human skin. *Yearbook of Physical Anthropology. American Journal of Physical Anthropology*. <https://doi.org/10.1002/ajpa.23737>

Quillen EE, Bauchet M, Bigham AW, Delgado-Burbano ME, Faust FX, Klimentidis YC, Mao X, Stoneking M, Shriver MD (2012) OPRM1 and EGFR contribute to skin pigmentation differences between indigenous Americans and Europeans. *Hum Genet*. 2012, 131: 1073-1080. 10.1007/s00439-011-1135-1.

Raghavan M, Steinrücken M, Harris K, et al. (2015) POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. 2015;349(6250):aab3884.

Raposo, G., & Marks, M. S. (2007). Melanosomes – dark organelles enlighten endosomal membrane transport. *Nature Reviews. Molecular Cell Biology*, 8(10), 786–797. <http://doi.org/10.1038/nrm2258>

Reales, G.; Rovaris, D.L.; Jacovas, V.C.; Hünemeier, T.; Sandoval, J.R.; Salazar-Granara, A.; Demarchi, D.A.; Tarazona-Santos, E.; Felkl, A.B.; Serafini, M.A.; et al. (2017) A tale of agriculturalists and hunter-gatherers: Exploring the thrifty genotype hypothesis in native South Americans. *Am. J. Physiol. Anthropol.* 2017, 163, 591–601.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, et al. (2012) Reconstructing Native American population history. *Nature* 488: 370–374.

Reimers, L. L., Crew, K. D., Bradshaw, P. T., Santella, R. M., Steck, S. E., Sirosh, I., ... Gammon, M. D. (2014). Vitamin D-related gene polymorphisms, plasma 25-hydroxyvitamin D, and breast cancer risk. *Cancer causes & control : CCC*, 26(2), 187–203. doi:10.1007/s10552-014-0497-9

Relethford, J.H. (2002) Apportionment of global human genetic diversity based on craniometrics and skin color. *Am. J. Phys. Anthropol.*, 118, 393–398.

Richards M.P., Schulting R.J., Hedges R.E. (2003) Sharp shift in diet at onset of Neolithic. *Nature*, 425 (2003), p. 366

Ritchie, M. D., & Van Steen, K. (2018). The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. *Annals of translational medicine*, 6(8), 157. doi:10.21037/atm.2018.04.05

Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F., & Moore, J. H. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *American journal of human genetics*, 69(1), 138–147. doi:10.1086/321276

Robins, Ashley H. (1991) *Biological perspectives on human pigmentation*. Cambridge University Press.

Rogers AR, Iltis D, Wooding S (2004) Genetic variation at the MC1R locus and the time since loss of human body hair. *Curr Anthropol.*, 45: 105-108. 10.1086/381006.

Romano R.A. Li H.X. Tummala R. Maul R. Sinha S. (2004) Identification of Basonuclin2, a DNA-binding zinc-finger protein expressed in germ tissues and skin keratinocytes *Genomics* 83 821 833

Ruxton G.D., Wilkinson D.M. (2011) Avoidance of overheating and selection for both hair loss and bipedality in hominins. *Proc. Natl. Acad. Sci. USA*, 108, pp. 20965-20969

Salzano FM e Bortolini MC (2002). *Genetics and Evolution of Latin American Populations*. Cambridge: Cambridge University Press

Santoro D, Lucisano S, Gagliostro G, Alibrandi A, Benvenega S, *et al.* (2015) Vitamin D receptor polymorphism in chronic kidney disease patients with complicated cardiovascular disease. *J Ren Nutr.*; 25:187-193. Ref.: <https://goo.gl/JDdtPo>

Scaglione F. & Panzavolta G. (2014) Folate, folic acid and 5-methyltetrahydrofolate are not the same thing. *Xenobiotica* 44 (5), 480–8.

Scherer, D.; Kumar, (2010) R. Genetics of pigmentation in skin cancer—A review. *Mutat. Res*, 705, 141–153.

Sedky NK, Abdel Rahman MF, Hassanein SI, Gad MZ.(2018) Genetic Variants of CYP2R1 Are Key Regulators of Serum Vitamin D Levels and

Incidence of Myocardial Infarction in Middle-Aged Egyptians. *Curr Pharm Biotechnol.* 2018;19(3):265-273. doi: 10.2174/1389201019666180528082737.

Sellers EA, *et al.* (2003) Adaptation of Inuit children to a low-calcium diet. *CMAJ.* 2003;168(9):1141–3.

Skoglund P, Mathieson I (2018) Ancient human genomics: The first decade. *Annu Rev Genomics Hum Genet* 19:8.1–8.24

Steckel RH, Rose JC (eds.). 2002 . *The backbone of history: Health and Nutrition in the Western Hemisphere.* Cambridge University Press.

Stoneking M. (2017) An Introduction to Molecular Anthropology. Published by John Wiley & Sons, Inc., Hoboken, New Jersey. ISBN 978-1-118-06162-6

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

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

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Richards O, Tyler-Smith C, Pena SD e Santos FR (2001) Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 68:1485–1496

Tiosano, D., Audi, L., Climer, S., Zhang, W., Templeton, A. R., Fernández-Cancio, M., ... Hochberg, Z. (2016). Latitudinal Clines of the Human Vitamin D Receptor and Skin Color Genes. *G3: Genes|Genomes|Genetics*, 6(5), 1251–1266. <http://doi.org/10.1534/g3.115.026773>

Veldurthy V , Wei R , Oz L , *et al* . (2016) Vitamin D, calcium homeostasis and aging. *Bone Res*;4:16041.doi:10.1038/boneres.2016.41

Visser M, Palstra RJ, Kayser M (2014) Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby BNC2 pigmentation gene. *Hum Mol Genet.* doi:10.1093/hmg/ddu289

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

Wainschtein P., Jain D. P., Yengo L., Zheng Z., *et al.* (2019) Recovery of trait heritability from whole genome sequence data. *bioRxiv* 588020; doi: <https://doi.org/10.1101/588020>

Walter H. (1971) Remarks on the environmental adaptation of man. *Humangenetik* 13:85–97

Wan X, Yang C, Yang Q, Xue H, Fan X, Tang N, *et al.* (2010) Boost: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet.* 2010; 87:325–40.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A (2007). Genetic variation and population structure in native Americans. *PLoS Genet.* 3(11):e185.

Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, *et al.* (2010) Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet.* 2010;376(9736):1808.

Wei, W-H, Hemani, G, Haley, CS (2014) Detecting epistasis in human complex traits. *Nature Reviews Genetics*, 15, 722– 733

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358. doi:10.2307/2408641

Wheeler PE (1992) The influence of the loss of functional body hair on the water budgets of early hominids. *J Hum Evol* 23: 379–388.

Wright, Sewall. (1951) The genetical structure of populations. *Annals of Eugenics.* 1951;15:323–354

Yanovich, R., Friedman, E., Milgrom, R., Oberman, B., Freedman, L., & Moran, D. S. (2012). Candidate gene analysis in Israeli soldiers with stress fractures. *Journal of sports science & medicine*, 11(1), 147–155.

Yao, S., Sucheston, L. E., Smiley, S. L., Davis, W., Conroy, J. M., Nowak, N. J., Hahn, T. (2011). Common genetic variants are associated with accelerated bone mineral density loss after hematopoietic cell transplantation. *PloS one*, 6(10), e25940. doi:10.1371/journal.pone.0025940

Zhang Z, He JW, Fu WZ, Zhang CQ, Zhang ZL (2013) An analysis of the association between the vitamin D pathway and serum 25-hydroxyvitamin D levels in a healthy Chinese population. *J Bone Miner Res.* Available: <http://www.ncbi.nlm.nih.gov/pubmed/23505139>. Accessed 25 March 2013.

Zhang, Y. & Liu, J. S. (2007). Bayesian inference of epistatic interactions in case–control studies. *Nature Genet.* 39, 1167–1173

Imagens Suplementares:

Figura suplementar 1

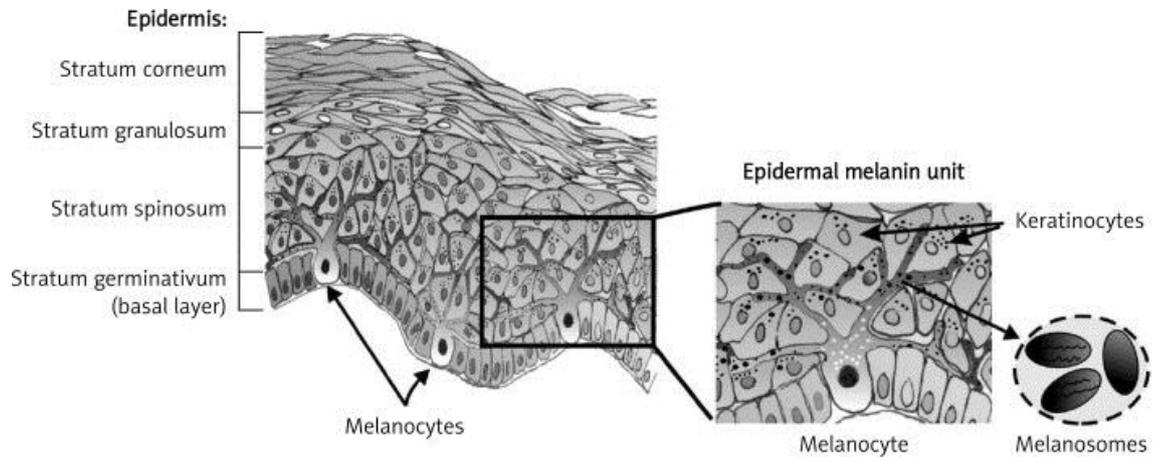


Figura suplementar 1: Esquema da estrutura da epiderme (copiado de Cichorek *et al.*, 2013).

Tabela S1. Classificação das populações estudadas.

	Grupo populacional:	Populações:	n° de indivíduos:	Família linguística*	Local da amostragem	Latitude:	Longitude:	Altitude:	Dieta:
	Inuits								
		EastGreenland	6	Eskimo-Aleut	Groenlândia	67.5	-37.9	< 2,500 m	Caçadores-coletores
		WestGreenland	4	Eskimo-Aleut	Groenlândia	65.3	-52	< 2,500 m	Caçadores-coletores
Soma		2	10						
	Mesoamericanos	Populações:							
		Kaqchikel	12	Northern Amerind	Guatemala	15	-91	< 2,500 m	Agriculturalistas
		Maya1	36	Northern Amerind	México	20.3	-87.8	< 2,500 m	Agriculturalistas
		Maya2	12	Northern Amerind	México	19.6	-89.7	< 2,500 m	Agriculturalistas
		Mixe	17	Northern Amerind	México	17	-96	< 2,500 m	Agriculturalistas
		Mixtec	5	Central-Amerind	México	17	-97	< 2,500 m	Agriculturalistas
		Purepecha	1	Chibchan-Paezan	México	19	-101.5	< 2,500 m	Agriculturalistas

		Tepehuano	25	Central- Amerind	México	23.2	-104.5	< 2,500 m	Agriculturalistas
		Zapotec1	22	Central- Amerind	México	16.5(16)	-97.2(-97)	< 2,500 m	Agriculturalistas
		Zapotec2	21	Central- Amerind	México	17.4	-96.7	< 2,500 m	Agriculturalistas
Soma		9	151						
	Andinos	Populações:							
		Aymara	23	Andean	Bolívia & Chile	-16.5(- 22)	-68.2(-70)	≥ 4,000 m	Agriculturalistas
		Quechua	40	Andean	Bolívia & Peru	-14.5(- 14)	-69(-74)	≥ 4,000 m	Agriculturalistas
		Yaghan	4	Andean	Chile	-55	-68	< 2,500 m	Caçadores- coletores
Soma		3	67						
	Amazônicos	Populações:							
		Arara	1	Ge-Pano- Carib	Brasil	-4	-53.5	< 2,500 m	Caçadores- coletores
		Arhuaco	5	Chibchan- Paezan	Colômbia	11	-73.8	< 2,500 m	Caçadores- coletores
		Guahibo	6	Equatorial- Tucanoan	Colômbia	5.8	-69.5	< 2,500 m	Caçadores- coletores
		Guarani	6	Equatorial- Tucanoan	Paraguai & Argentina	-23(- 22.5)	-54(-63.8)	< 2,500 m	Caçadores- coletores

		Jamamadi	1	Equatorial-Tucanoan	Brasil	-8.5	-64.5	< 2,500 m	Caçadores-coletores
		Kaingang	1	Ge-Pano-Carib	Brasil	-24	-52.5	< 2,500 m	Caçadores-coletores
		Karitiana	13	Equatorial-Tucanoan	Brasil	-10	-63	< 2,500 m	Caçadores-coletores
		Kogi	4	Chibchan-Paezan	Colômbia	11	-74	< 2,500 m	Caçadores-coletores
		Maleku	3	Chibchan-Paezan	Costa Rica	10.6	-84.8	< 2,500 m	Caçadores-coletores
		Palikur	3	Equatorial-Tucanoan	Guiana	4	-51.8	< 2,500 m	Caçadores-coletores
		Parakana	1	Equatorial-Tucanoan	Brasil	-4.8	-50	< 2,500 m	Caçadores-coletores
		Piapoco	7	Equatorial-Tucanoan	Colômbia	3	-68	< 2,500 m	Caçadores-coletores
		Surui	24	Equatorial-Tucanoan	Brasil	-11	-62	< 2,500 m	Caçadores-coletores
		Teribe	3	Chibchan-Paezan	Costa Rica	9	-83.2	< 2,500 m	Caçadores-coletores
		Ticuna	6	Equatorial-Tucanoan	Colômbia	-3.81	-70.01	< 2,500 m	Caçadores-coletores
		Toba	4	Ge-Pano-Carib	Argentina	-26.5	-59.3	< 2,500 m	Caçadores-coletores

		Waunana	3	Chibchan-Paezan	Colômbia	5	-77	< 2,500 m	Caçadores-coletores
		Wayuu	9	Equatorial-Tucanoan	Colômbia	11	-73	< 2,500 m	Caçadores-coletores
		Wichi	5	Ge-Pano-Carib	Argentina	-22.5	-63.8	< 2,500 m	Caçadores-coletores
Soma		19	105						

*De acordo com Greenberg

Tabela S2. Genes envolvidos no metabolismo da vitamina D

Símbolo	Nome	Localização Genômica/ Banda Citogenética	Função
CYP27B1	Membro 1 da subfamília B da família 27 do Citocromo P450	chr12:57,762,334-57,768,986 (GRCh38/hg38) 12q14.1	A proteína codificada por este gene se localiza no interior da membrana mitocondrial onde hidroxila a 25-hidroxitamina D ₃ na posição 1 α . Esta reação sintetiza a 1 α ,25-dihidroxitamina D ₃ (1,25[OH]2D ₃), que é a forma ativa da vitamina D ₃ que se liga ao receptor de vitamina D e regula o metabolismo do cálcio (Christakos, et al., 2016).
CUBN	Cubilina	chr10:16,823,964-17,130,249 (GRCh38/hg38) 10p13	A Cubilina é um receptor de superfície para a proteína DBP no túbulo proximal do rim. Esta proteína se associa à Megalina para internalizar o complexo DBP-25(OH) D ₃ . Ela tem como função reabsorver proteínas que seriam eliminadas na urina (Christakos, et al., 2016).
CYP24A1	Membro 1 da subfamília A da família 24 do Citocromo P450	chr20:54,153,449-54,173,985 (GRCh38/hg38) 20q13.2	Este gene codifica uma proteína mitocondrial que degrada a forma ativa da vitamina D ₃ , (1,25[OH]2D ₃), por hidroxilação da cadeia lateral. Assim, a proteína regula os níveis de vitamina D ₃ , desempenhando um papel na homeostase do cálcio e no sistema endócrino da vitamina D (Christakos, et al., 2016).
CYP2R1	Membro 1 da subfamília R da família 2 do Citocromo P450	chr11:14,877,436-14,894,034 (GRCh38/hg38) 11p15.2	A enzima codificada por este gene é uma hidroxilase microssomal que converte a vitamina D em calcidiol, a 25(OH)D ₃ . Este processo ocorre por meio da hidroxilação do carbono 25 do colecalciferol (Hosseini-nezhad e Holick, 2013; Christakos, et al., 2016).

LRP2	Proteína relacionada ao receptor de lipoproteínas de baixa densidade	2 chr2:169,127,109-169,362,685 (GRCh38/hg38) 2q31.1	O gene LRP2 codifica a proteína Megalina, que é uma glicoproteína receptora com múltiplos ligantes. Dentre outras funções, ela tem papel na sinalização celular e é importante para a reabsorção de vários metabólitos, tais como esteroides, proteínas de ligação à vitaminas e hormônios.
GC	Proteína de ligação vitamina D	chr4:71,741,693-71,805,520 (GRCh38/hg38) 4q13.3	A proteína codificada por este gene pertence à família de genes da albumina. É uma proteína multifuncional que se liga à vitamina D e seus metabólitos plasmáticos e os transporta para seus tecidos alvo.

Os dados foram coletados dos seguintes sites: Ensemble (<http://www.ensembl.org/>); OMIM (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>); Uniprot (<http://beta.uniprot.org/>); Genecards (<http://www.genecards.org/index.shtml>); GHR (<https://ghr.nlm.nih.gov/>). As indicações de bibliografia complementar estão citadas entre parênteses e as referências completas encontram-se no item Referências Bibliográficas.

Tabela S3. Genes envolvidos no controle da expressão da vitamina D

Símbolo	Nome	Localização Genômica/ Banda Citogenética	Função
RXRA	Receptor retinoide X subtipo alfa	chr9:134,317,098-134,440,586 (GRCh38/hg38) 9q34.2	Os receptores retinoides X (RXRs) e os receptores de ácido retinóico (RARs) são receptores nucleares que medeiam os efeitos biológicos dos retinóides por meio do envolvimento na ativação gênica mediada por ácido retinóico (metabólito ativo da vitamina A). Esses receptores funcionam como fatores de transcrição pela ligação, como heterodímeros ou homodímeros, a sequencias específicas em promotores de gens alvo. A proteína

codificada por este gene é membro da superfamília de receptores de hormônio que regulam a transcrição. Os RXRs são membros da família de receptores nucleares NR2B e se ligam a outros receptores nucleares como os PPARs, LXRs e VDRs.

VDR	Receptor de vitamina D	chr12:47,841,537-47,943,048 (GRCh38/hg38) 12q13.11	Este gene codifica um receptor nuclear de hormônio para vitamina D. Ele é um fator de transcrição transatruante. Sua sequência é semelhante a receptores de hormônios da tireóide e esteróides. Alvos à jusante (<i>downstream</i>) deste receptor nuclear de hormônio estão envolvidos principalmente no metabolismo mineral, embora o VDR regule uma série de outras vias metabólicas, como as envolvidas na resposta imune e câncer. Mutações neste gene estão associadas ao raquitismo dependente de vitamina D tipo II. O encadeamento alternativo resulta em múltiplas variantes transcricionais, codificando diferentes proteínas. Este fator de transcrição medeia a ação da vitamina D, controlando a expressão de genes sensíveis ao hormônio. Os receptores de vitamina D (VDRs) são membros da família NR1I, eles formam heterodímeros com os membros da família do receptor retinoide X.
------------	------------------------	---	---

Os dados foram coletados dos seguintes sites: Ensemble (<http://www.ensembl.org/>); OMIM (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>); Uniprot (<http://beta.uniprot.org/>); Genecards (<http://www.genecards.org/index.shtml>); GHR (<https://ghr.nlm.nih.gov/>). As indicações de bibliografia complementar estão citadas entre parênteses e as referências completas encontram-se no item Referências Bibliográficas.

Tabela S4. Genes relacionados com pigmentação de pele:

Símbolo	Nome	Localização Genômica/ Banda Citogenética	Função
<i>BNC2</i>	Basonuclina 2	chr9:16,409,503- 16,870,843 (GRCh38/hg38) 9p22.3-p22.2	Codifica uma proteína pertencente a família de fatores de transcrição Basonuclina, com motivos dedo de zinco que se ligam a sequências de DNA. A Basonuclina 2 está envolvida no processamento do mRNA e pode atuar como fator de transcrição. Atua na pigmentação através da expressão nos melanócitos, controlando a saturação da cor de pele. Mutações neste gene estão associadas a manchas faciais (Romano, <i>et al.</i> , 2004; Visser <i>et al.</i> , 2014).
<i>DCT</i>	Dopacroma Tautomerase	chr13:94,436,808- 94,550,265 (GRCh38/hg38) 13q32.1	Este gene codifica uma proteína que catalisa a conversão do L-dopacroma em ácido 5,6-dihidroxi-indol-2-carboxílico (DHICA). Ele tem expressão restrita a pele e está envolvido na regulação dos níveis de eumelanina e feomelanina.
<i>EGFR</i>	Receptor do fator de crescimento epidérmico	chr7:55,019,021- 55,211,628 (GRCh38/hg38) 7p11.2	A proteína codificada por este gene é uma glicoproteína transmembrana membro da superfamília de proteínas quinase. É uma proteína receptora de superfície celular que se liga ao fator de crescimento epidérmico. Quando ligada induz a dimerização do receptor e a auto fosforilação da tirosina, conduzindo a proliferação celular. Este receptor atua na regulação dos queratinócitos, que influenciam na quantidade de melanócitos (Quillen, <i>et al.</i> , 2012).

HERC2	Domínios HECT e RLD contendo Ubiquitina proteína ligase E3 2	chr15:28,111,037-28,322,172 (GRCh38/hg38) 15q13.1	Este gene pertence a família HERC que codifica um grupo de proteínas normalmente grandes, com múltiplos domínios estruturais. A proteína codificada por ele tem função associada a retenção de proteínas de reparação cromossômica via ligação de ubiquitina. Variações neste gene estão associadas a diferenças na cor de pele, olhos e cabelos. O SNP associado a cor neste gene encontra-se em uma região intrônica que regula a expressão do OCA2 (Visser, <i>et al.</i> , 2012).
KITLG	Ligante de KIT	chr12:88,492,793-88,580,851 (GRCh38/hg38) 12q21.32	Codifica o receptor de tirosina-quinase pelo locus KIT. Acredita-se que seja um fator pleiotrópico que atua na migração celular. Dentre outras funções, desempenha um papel essencial na regulação da proliferação e sobrevivência de muitos tipos celulares, incluindo melanócitos, células sanguíneas e células germinativas (Guenther <i>et al.</i> , 2014).
MC1R	Receptor de melanocortina 1	chr16:89,912,119-89,920,977 (GRCh38/hg38) 16q24.3	Codifica um receptor acoplado à proteína G com sete hélices transmembranares na superfície dos melanócitos. Esta proteína é receptora para os hormônios estimulantes de melanócitos (MSH) e ativa a cascata de cAMP. Ela controla o tipo de melanina produzida nos melanócitos. Quando o receptor está ativo desencadeia uma cascata de reações químicas que levam à produção de eumelanina, quando inativo, ocorre a produção de feomelanina.
MFSD12	Proteína que contém domínio da superfamília de facilitadores	12 chr19:3,538,261-3,574,290 (GRCh38/hg38) 19p13.3	O MFSD12 codifica uma proteína nuclear que provavelmente desempenha um papel na proliferação da regulação celular. Este gene é expresso nos rins, na pele e em 25 outros tecidos. Acredita-se que ele

	principais		influencia a produção e armazenamento da melanina na pele. Análises funcionais indicaram que este gene pode alterar a coloração de animais modelo. Um SNP neste gene ou próximo a ele está fortemente associado a pigmentação de pele em populações africanas e outra variante se relaciona com cor de pele em asiáticos e nativos americanos. (Crawford, et al., 2017; Adhikari et al., 2019).
OCA2	Proteína transmembrana melanossomal OCA2	chr15:27,754,873-28,099,358 (GRCh38/hg38) 15q13.1	A proteína de membrana codificada por este gene pode estar envolvida no transporte da tirosina dentro dos melanócitos. Ela regula o pH e a maturação dos melanossomos. Parece regular o processamento pós-traducional da tirosinase. O splicing alternativo neste gene resulta múltiplas variações no transcrito. Este gene está envolvido na determinação da cor de pele e acredita-se que seja o principal determinante da cor dos olhos. Mutações em tal gene podem resultar em albinismo oculocutâneo tipo 2.
OPRM1	Receptor opioide Mu 1	chr6:154,010,496-154,246,867 (GRCh38/hg38) 6q25.2	Este gene codifica um dos receptores de opioides humanos, o receptor Mu (MOR). Ele é o principal alvo para peptídeos opioides endógenos e agentes analgésicos opioides como a beta-endorfina e as encefalinas. O MOR também desempenha um importante papel na dependência a drogas de abuso através da modulação do sistema da dopamina. Um alelo neste gene foi associado a diferença na pigmentação de pele entre europeus e nativos americanos. Os receptores de opioides são ativos em queratinócitos e melanócitos e podem influenciar a diferenciação dessas células

			(Quillen, <i>et al.</i> , 2012). Também existe a possibilidade de a influência do OPRM1 na pigmentação ser mediada pelo comportamento de busca pelo sol (Khouja <i>et al.</i> , 2018).
SLC24A5	Membro 5 da família 24 portadores soluto	chr15:48,120,972-48,142,672 (GRCh38/hg38) 15q21.1	Este gene é membro da família de permutadores de sódio/cálcio dependente de potássio. Ele codifica uma proteína de membrana intracelular com dois conjuntos de múltiplos segmentos transmembranares e dois grandes loops hidrofílicos. A proteína codificada por ele provavelmente está envolvida no transporte de íons nos melanossomos. Diferenças na sequência desse gene estão associadas a variações na cor de pele.
SLC45A2	Membro 2 da família 45 portadores soluto	chr5:33,944,616-33,985,144 (GRCh38/hg38) 5p13.2	Este gene codifica uma proteína transportadora que medeia a síntese de melanina. Esta proteína é bastante expressa em linhagens celulares de melanoma. Mutações neste gene são a causa de albinismo oculocutâneo tipo 4, e polimorfismos nele estão associados com variações na cor de pele e cabelo.
TYR	Tirosinase	chr11:89,177,452-89,295,759 (GRCh38/hg38) 11q14.3	A enzima codificada por este gene catalisa os primeiros dois passos e pelo menos um passo subsequente da conversão da tirosina para melanina. A enzima possui atividade catalítica de tirosina hidroxilase e dopa oxidase e requer cobre para funcionar. Polimorfismos neste gene resultam em variações na pigmentação de pele.

<i>TYRP1</i>	Proteína relacionada à Tirosinase	1 chr9:12,685,439-12,710,290 (GRCh38/hg38) 9p23	Este gene codifica uma enzima melanossomal que pertence à família da tirosinase e desempenha um papel importante na via de biossíntese da melanina. A enzima catalisa a oxidação de moléculas específicas, influenciando o tipo de melanina que será sintetizada. Em menor grau é também capaz de hidroxilar a tirosina e produzir melanina.
---------------------	-----------------------------------	--	--

Os dados foram coletados dos seguintes sites: Ensemble (<http://www.ensembl.org/>); OMIM (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>); Uniprot (<http://beta.uniprot.org/>); Genecards (<http://www.genecards.org/index.shtml>); GHR (<https://ghr.nlm.nih.gov/>). As indicações de bibliografia complementar estão citadas entre parênteses e as referências completas encontram-se no item Referências Bibliográficas.

As tabelas suplementares 5 e 6 estão disponíveis para *download* no seguinte link:

<https://drive.google.com/drive/folders/1mXVPoEI68ilm16d3oCnLnYzyY24qzBb?usp=sharing>

Tabela S7: Redes identificadas usando o limiar CCC ≥ 0.65 com as amostras descritas na Tabela S1 e os SNPs descritos na Tabela S8 (arquivo 'S.8_Nets_interações.xlsx')

Rede	Gene	SNP	Alelo	Cromossomo
R_1_LRP2	LRP2	rs2302694	C	2
	LRP2	rs2302693	A	2
	LRP2	rs9789747	T	2
	LRP2	rs3926693	G	2
	LRP2	rs830994	G	2
R_2_OPRM1	OPRM1	rs1294092	A	6
	OPRM1	rs518596	C	6
	OPRM1	rs6557337	A	6
R_3_EGFR	EGFR	rs10488140	C	7
	EGFR	rs11977660	C	7
	EGFR	rs11773818	T	7
	EGFR	rs2058502	T	7
R_4_CUBN_OCA2 _EGFR_TYRP1	CUBN	rs10752062	C	10
	CUBN	rs3740164	C	10
	OCA2	rs7171012	C	15
	OCA2	rs7495174	A	15
	EGFR	rs2293347	G	7
	TYRP1	rs2762462	T	9
	CUBN	rs780807	T	10
	CUBN	rs1996316	G	10
	CUBN	rs10795445	T	10
	CUBN	rs2281649	G	10
	CUBN	rs7916688	G	10
	OCA2	rs750358	T	15
R_5_TYR	TYR	rs1939258	C	11
	TYR	rs3793975	C	11
	TYR	rs621313	T	11
	TYR	rs1432447	T	11
	TYR	rs11018541	G	11
R_6_VDR_CYP24A1	VDR	rs2107301	C	12
	CYP24A1	rs2248359	C	20
	VDR	rs11574143	G	12
	VDR	rs4516035	T	12
	CYP24A1	rs3787557	T	20
R_7_OCA2	OCA2	rs3947367	A	15

	OCA2	rs12898424	T	15
R_8_OCA2	OCA2	rs11852452	C	15
	OCA2	rs4778220	A	15
R_9_LRP2	LRP2	rs831022	T	2
	LRP2	rs16856843	G	2

A tabela suplementar 8 está disponível para *download* no seguinte link:

<https://drive.google.com/drive/folders/1mXVPoEI68ilm16d3oCnLnYzyY24qzBb?usp=sharing>