UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

DESVENDANDO MECANISMOS GENÉTICOS DE ADAPTAÇÃO À ALTA ALTITUDE EM POPULAÇÕES NATIVAS AMERICANAS DOS ANDES

Vanessa Cristina Jacovas
Orientadora: Prof.ª Dr.ª Maria Cátira Bortolini

Tese submetida ao Programa de Pós-graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito para obtenção do título de Doutor em Ciências.

Porto Alegre
Março de 2019.

Este trabalho foi desenvolvido no Laboratório de Evolução Humana e Molecular da Universidade Federal do Rio Grande do Sul e no Laboratório de Diversidade Humana, no Departamento de Ciências Experimentais e da Saúde da Universidade Pompeo Fabra, no Parque de Pesquisa Biomédica de Barcelona, entre março de 2016 e fevereiro de 2017, com o financiamento do Conselho Nacional de Pesquisa (CNPq).

AGRADECIMENTOS

Aos voluntários Nativos Americanos que aceitaram contribuir para que estudos científicos dessa natureza pudessem ser realizados.

À minha orientadora, Prof.ªDr.ª Maria Cátira Bortolini, pela confiança em mim depositada, pelos conhecimentos transmitidos diariamente, pela orientação, incentivo e amizade.

In memoriam, ao Prof. Dr. Francisco Mauro Salzano, pelo exemplo de pesquisador e cientista, pelas eternas palavras de otimismo. Professor, quanto mais nós conhecemos o mundo, mais te admiramos.

Ao Prof. Dr. David Comas pela atenção, confiança e por me receber tão bem em seu grupo de pesquisa.

À Prof.ª Dr.ªTábita Hünemeier, por abrir as portas do seu laboratório na Universidade de São Paulo e me proporcionar uma enorme oportunidade de crescimento.

Aos meus grandes colegas do Laboratório de Evolução Humana e Molecular pelo excelente ambiente de trabalho, por todos os cafés, almoços, e pelas palavras de incentivo nos dias mais difíceis.

Aos colegas com quem tive o prazer de trabalhar na Universidade Pompeu Fabra.

Ao meu companheiro, Rafael, pela amizade e parceria, pela compreensão em todos os momentos, pelas palavras de apoio de todos os dias.

Por último, mas não menos importante, à minha família, pelo amor e apoio incondicional em todos os momentos da minha vida. Por todo o esforço que vocês fizeram pra que nós, eu e o mano, tivéssemos a oportunidade se ser quem nós quiséssemos ser.

LISTA DE ABREVIATURAS

ABCA1	*atp-binding cassette subfamily A, member 1*
ANGPT1	*angiopoietin 1*
ARHGAP15	*rho GTPase activating protein 15*
ARNT	*aryl hydrocarbon receptor nuclear translocator*
atm	atmosfera
ATP	adenosina trifosfato
BHLHE41	*basic helix-loop-helix family member e41*
BRINP3	*bone morphogenetic protein/retinoic acid inducible neural-specific 3*
C6orf195	*chromosome 6 open reading frame 195*
CBARA1	*calcium-binding atopy-related autoantigen 1*
CCDC141	*coiled-coil domain containing 141*
COL6A1	*collagen type 6 alpha-1*
CTBP2	*c-terminal binding protein 1*
ECE1	*endothelin converting enzyme 1*
EDRF1	*erythroid differentiation regulatory factor 1*
eNOS	sintase endotelial do óxido nítrico
EPAS1	*endothelial PAS domain-containing protein 1*
EPO	*erythropoietin*
FADS	*fatty acid desaturases*
FAM213A	*family with sequence similarity 213, member A*
FOXO1	*forkhead box O1*
FVC	capacidade vital forçada
GWAS	estudos de associação com varredura genômica
HGF	*hepatocyte growth factor*
HIF	fator de indução por hipóxia
HLA-DR	*human leukocyte antigen-DR alpha*
HLA-G	antígeno leucocitário humano-G
HLH	*helix-loop-helix*
HMOX2	*heme oxygenase 2*
HREs	elementos responsivos à hipóxia

kb	kilobases
KCTD12	*k+ channel tetramerization domain containing 12*
LEPR	*leptin receptor*
LGM	último máximo glacial
LINE-1	*long interspersed element-1*
m	metros
MDM2	*mouse double minute 2 homolog*
MHC	complexo principal de histocompatibilidade
mmHg	milímetro de mercúrio
MTHFD1	*methylenetetrahydrofolate dehydrogenase, cyclohydrolase and formyltetrahydrofolate synthetase 1*
MTHFR	*methylenetetrahydrofolate reductase*
NO	óxido nítrico
NOS2	nitric oxide synthase 2
NOS3	nitric oxid synthase 3
O2	oxigênio
ºC	graus Celsius
PAPOLA	*poly(A) polymerase alpha*
PAS	*per arnt sim*
PHD	*hypoxia-inducible factor prolyl hydroxylase 1*
PHD2	*hypoxia-inducible factor prolyl hydroxylase 2*
PPARA	*peroxisome proliferator-activated receptor A*
PRKAA1	*protein kinase AMP-activated catalytic subunit alpha 1*
PTGIS	*prostaglandin I2 synthase*
RCIU	restrição do crescimento intrauterino
RNF216	*ring finger protein 216*
RORA	*RAR related orphan receptor A*
RUNX1	*runt related transcription factor 1*
RUV	radiação unltravioleta
RXR	*receptor retinóide X*
RYR1	*ryanodine receptor 1*
SaO2	saturação de oxigênio

SFTPD	*surfactant protein D*
sHLA-G	HLA-G solúvel
SLC30A9	*solute carrier family 30 member 9*
SNP	polimorfismo de nucleotídeo único
STRs	*short tandem repeats*
TBX5	T-Box 5
TED	*Tibetan enriched deletion*
TEX36	*testis expressed 36*
THRB	*thyroid hormone Receptor beta*
TP53	*tumor protein P53*
VAV3	*vav guanine nucleotide exchange factor 3*
VDR	*vitamin D3 receptor*
VEGF	fator de crescimento endotelial vascular
VEGFA	*vascular endothelial growth factor A*
VHL	*von Hipper-Lindau*
VRK1	*vaccinia related kinase 1*

LISTA DE FIGURAS E TABELAS

SUMÁRIO

RESUMO

O estudo da adaptação humana para vida em altas altitudes é um tema bastante abordado na literatura científica mundial, tanto no que se refere às adaptações genéticas, quanto às adaptações fisiológicas. Historicamente, os estudos nesta área iniciaram-se já nos anos 1900, quando ainda se acreditava que os nativos dessas terras eram seres fisicamente inferiores, sendo que funções fisiológicas "normais" eram consideradas impossíveis em altitudes onde o oxigênio é encontrado em concentrações tão baixas. Essa perspectiva começou a ser alterada quase na década de 50, quando cientistas descreveram a existência de mecanismos fisiológicos que, ao longo dos séculos, aclimataram essas populações ao ambiente extremo ao qual vivem.

No final da década de 70 (início dos anos 80) iniciaram-se os primeiros estudos com adaptação genética para a vida em altas altitudes. Após quase 4 décadas de investigação e mais de duas centenas de publicações relativas à adaptação genética às altas altitudes, ainda não existe um consenso sobre qual gene, ou genes, estão principalmente envolvidos nesse fenômeno.

Deste modo, as abordagens apresentadas nesta tese buscam preencher as lacunas encontradas nesta temática, principalmente no que tange às populações do Altiplano Andino, ainda insuficientemente estudadas e compreendidas. Os resultados desta tese foram apresentados sob a forma de 3 artigos científicos, e os resultados são resumidos como seguem:

Jacovas *et al*., 2015: A diversidade de cinco polimorfismos de nucleotídeo único (SNPs) localizados nos genes da via *TP53* (*TP53*, rs1042522; *MDM2*, rs2279744; *MDM4*, rs1563828; *USP7*, rs1529916; e *LIF*, rs929271) foram estudadas em um total de 282 indivíduos pertencentes às seguintes populações: Quechua, Aymara, Chivay, Cabanaconde, Yanke, Taquile, Amantani, Anapia, Uros, Guarani Ñandeva e Guarani Kaiowá, caracterizadas como Nativas Americanas apresentando alto nível (> 90%) de ancestralidade indígena. Além disso, acrescentamos ao nosso banco de dados originais, dados já publicados referentes a 100 indivíduos de outras cinco populações Nativas Americanas (Suruí, Karitiana, Maya, Pima e Piapoco). As populações foram classificadas como vivendo em altas altitudes [≥ 2.500 metros (m)] ou em terras baixas (<2.500 m). Nossas análises revelaram que os alelos *USP7*-G, *LIF*-T e *MDM2*-T mostraram evidências significativas de

que foram selecionados e correlacionados a variáveis ambientais adversas inerentes às altas altitudes. Nossos resultados mostram, pela primeira vez, que alelos da rede clássica de *TP53* foram evolutivamente cooptados para o sucesso da colonização humana dos Andes.

Jacovas *et al*., 2018: O Altiplano Andino tem sido ocupado continuamente desde o final do Pleistoceno, há cerca de 12.000 anos atrás, o que situa os nativos andinos como uma das populações mais antigas a viverem em altas altitudes. No presente estudo, analisamos dados genômicos de Nativos Americanos que vivem há muito tempo no Altiplano Andino e nas áreas de planície da Amazônia e Mesoamérica. Nós identificamos três novos genes candidatos - *SP100*, *DUOX2* e *CLC* - com evidência de seleção positiva para adaptação à altas altitudes em andinos. Esses genes estão envolvidos na via de *TP53* e relacionados a vias fisiológicas importantes para a resposta à hipóxia às altas altitudes, como aquelas ligadas ao aumento da angiogênese, adaptações do músculo esquelético e funções imunes na interface materno-fetal. Nossos resultados, combinados com outros estudos, mostraram que os Andinos se adaptaram ao Altiplano de diferentes maneiras e usando estratégias moleculares distintas em comparação com os de outros nativos que vivem em grandes altitudes.

Jacovas *et al*., a ser submetido para publicação: A resposta fisiológica à hipóxia envolve uma ampla gama de mecanismos biológicos e, nesse cenário, residentes de altas altitude são geneticamente adaptados. O HLA-G desempenha um papel fundamental em vários processos biológicos, incluindo a reprodução humana e a resposta à hipóxia, sendo um gene potencialmente envolvido na adaptação humana às altitudes elevadas. A variação genética do *HLA-G* já foi caracterizada e mostrou ter grande impacto nos padrões de expressão da proteína. Estudos em nível mundial de populações sugerem que a seleção balanceadora esteja atuando no *HLA-G* principalmente nas regiões regulatórias. No entanto, o conhecimento atual da variação do *HLA-G* nas populações Nativas Americanas e, especialmente, na região 3' UTR não traduzida, ainda é inexplorado. Além disso, a hipótese de que variações específicas nesta região contribuem para a resposta de hipóxia em tais populações ainda não havia sido investigado. Neste estudo, o *HLA-G* 3'UTR foi sequenciado em 301 indivíduos ameríndios de 17 populações (Amantani, Anapia, Andoas, Aimara, Cabanaconde, Chivay, Cinta Larga, Guarani Kaiowá, Guarani Ñandeva, Lengua, Quechua, Taquile, Uros, Xavante, Xicrin, Yanke e Zoró), divididos em residentes de altas altitudes (>2.500m) e habitantes de terras baixas (<2.500m). No total, 11 haplótipos foram

observados, três deles únicos de altas altitudes (UTR-8, -18 e -30) e outros três apenas em terras baixas (UTR-9, -13 e -21). Uma estrutura populacional foi observada nesses dois grupos populacionais e, curiosamente, a variância observada dentro dos grupos (*Fst*) é marcadamente alta (11%) para a UTR-5, indicando uma diferença notável e significativa entre as distribuições de haplótipos. Além disso, análises posteriores mostraram correlações estatísticas entre o UTR-5 e o UTR-2 com diversas variáveis climáticas, incluindo radiação ultravioleta e altitude.

Desta forma nossos dados corroboram a ideia de que não há somente uma única resposta adaptativa ao estresse devido às altas altitudes. Apesar disto, sugerimos que uma regulação adequada de p53 (e proteínas associadas), bem como de HLA-G esteja relacionada à adaptação de altas altitudes nas populações andinas, de forma que não apenas permita a sobrevivência dessas populações, mas também seu sucesso reprodutivo em ambientes hostis. Essa combinação de alelos, no entanto, pode resultar em suceptibilidade ao câncer nos dias atuais. Os mecanismos envolvidos na regulação destas vias, que tem em comum o "gatilho" da hipóxia, bem como suas consequências funcionais, merecem ser alvo de futuras investigações.

ABSTRACT

The study of human adaptation to life at high altitudes is a widely addressed topic in the world scientific literature with most of studies investigating both genetic and physiological adaptations. Historically, the investigations in this field began in the early 1900 when it was still believed that the native populations belonging to these lands were physically inferior individuals since normal physiological functions were considered insufficient at high altitudes where the oxygen concentrations are limited. This perspective began to change in the '50s when scientists described the existence of physiological mechanisms that over the centuries shaped these populations to the harsh environment of high altitude.

In the late '70s (early '80s) the first studies of genetic adaptation to life at high altitudes began. After almost four decades of investigation and more than two hundred publications relating to genetic adaptation to high altitudes, a consensus of which genetic variant or genes are involved in this phenomenon is still a matter of interest.

In this context, the thematic presented in this thesis seek to fill the gaps in the knowledge of high altitude genetic adaptation. More specifically, we investigated the populations of the Andean Altiplano which are still insufficiently studied and understood. The current thesis are presents in the form of 3 scientific articles, and the results are summarized as follow:

Jacovas *et al*., 2015: The diversity of five single nucleotide polymorphisms (SNPs) located in genes of the *TP53* pathway (*TP53*, rs1042522; *MDM2*, rs2279744; *MDM4*, rs1563828; *USP7*, rs1529916; and *LIF*, rs929271) were studied in a total of 282 individuals belonging to Quechua, Aymara, Chivay, Cabanaconde, Yanke, Taquile, Amantani, Anapia, Uros, Guarani Ñandeva, and Guarani Kaiowá populations. These individuals were characterized as Native American showing a high level (> 90%) of Native American ancestry. In addition, published data of 100 individuals from five Native American populations (Surui, Karitiana, Maya, Pima, and Piapoco) were analyzed. The populations were classified high altitude [≥ 2,500 meters (m)] or lowlands (< 2,500 m) populations. Our analyses revealed that alleles*USP7*-G, *LIF*-T, and *MDM2*-T showed significant evidence that they were selected along to the settlement of the harsh environmental variables related to high altitudes. Our results show for the first time that

alleles of classical *TP53* network genes have been evolutionary co-opted for the successful human colonization of the Andes.

Jacovas *et al*., 2018: The Andean Altiplano has been continuously settled since the late Pleistocene, ~12,000 years ago, a fact that places the Andean natives as one of the most ancient populations living at high altitudes. In the present study, we analyzed genomic data from Native Americans living a long-time at Andean Altiplano high altitude and Amazonia and Mesoamerica lowland areas. We have identified three new candidate genes - *SP100*, *DUOX2*, and *CLC* - with evidence of positive selection for altitude adaptation in Andeans. These genes are involved in the *TP53* pathway and are related to physiological routes important for high-altitude hypoxia response, such as those linked to increased angiogenesis, skeletal muscle adaptations, and immune functions at the fetus-maternal interface. Our results, combined with other studies, showed that Andeans have adapted to the Altiplano in different ways and using distinct molecular strategies as compared to those of other natives living at high altitudes.

Jacovas *et al*., To be submitted for publication. The physiological response to hypoxia involves a wide range of biological mechanisms. In this scenario, local residents of highaltitude areas are genetically adapted. HLA-G plays a crucial role in several biological processes, including human reproduction and hypoxia response, and thus being a gene potentially involved in human adaptation to high altitudes. The genetic variation of *HLA-G* is well known and it has been shown that regulatory SNPs have a significant impact on protein expression patterns. Studies evaluating worldwide populations suggest that balancing selection is acting in the *HLA-G* mainly in the regulatory regions. Interestingly, the *HLA-G* variation in Native American communities (living at low and high altitude), and more specifically the characterization of the whole 3 'untranslated region (UTR) is still unexplored. In addition, the hypothesis that functional SNPs in this region contribute to hypoxia response in such populations is of wide biological relevance. In this study, *HLA-G* 3'UTR was characterized in 301 Amerindian individuals from 17 populations (Amantani, Anapia, Andoas, Aimara, Cabanaconde, Chivay, Cinta Larga, Guarani Kaiowá, Guarani Ñandeva, Language, Quechua, Taquile, Uros, Xavante , Xicrin, Yanke and Zoró), grouped into residents of high altitudes (> 2,500 m) and inhabitants of low lands (<2,500 m). In total, 11 *HLA-G* 3'UTR haplotypes were observed, three of them exclusive to high altitudes (UTR-8, -18 and -30) and three others exclusive to lowlands

(UTR-9, -13 and -21). A population structure was observed in these two population groups and, interestingly, the observed variance within the groups (Fst) was markedly high (11%) for the UTR-5, indicating a significant difference between the haplotype distributions. Also, subsequent analyzes showed statistical correlations between UTR-5 and UTR-2 with several environmental variables, including ultraviolet radiation and altitude.

Therefore, our data corroborate the idea that there is not only a single adaptive response to stress due to high altitudes. Here, we suggest that adequate regulation of p53 (and associated proteins) as well as HLA-G is related to the high and long-time adaptation in the Andeas, ensuring not only the survival of the highlanders, but also their reproductive success in hostile environments. This combination of alleles, however, may result in cancer susceptibility at present days. The mechanisms involved in the regulation of these pathways, which have in common the "trigger" of hypoxia, as well as their functional consequences, deserve to be the subject of future investigations.

CAPÍTULO 1
INTRODUÇÃO

1. INTRODUÇÃO

1.1 Considerações gerais

Os estudos sobre adaptabilidade humana às grandes altitudes começaram no início dos anos 1900, quando pesquisadores franceses e britânicos lideraram expedições para as terras altas do México e do Peru. A maioria destes pesquisadores acredita que o nativo dessas terras era um ser fisicamente inferior, já que funções fisiológicas "normais" eram consideradas impossíveis em altitudes onde o oxigênio é encontrado em concentrações tão baixas (Cueto, 1989). Essa perspectiva foi abruptamente alterada quando, em 1927, Carlos Monge Medrano - um médico peruano que posteriormente se especializaria em medicina da altitude - organizou uma expedição às terras altas do Peru, onde confirmou a existência de mecanismos fisiológicos que ao longo dos séculos aclimataram essas populações à baixa pressão de oxigênio característica às grandes altitudes. Diz ele numa publicação anos mais tarde: "*Biologically we have shown that the Andean carries in his organism the hereditary and ancestral soma which permits life at the very great altitudes that mark certain large inhabited areas of South America*" (Monge, 1948).

Ao longo de muitas gerações, os povos que vivem nas regiões de altas altitudes [acima de 2.500 m (metros) de acordo com Moore 2001] do globo terrestre (Altiplano Andino, na América do Sul, Planalto do Tibete, na Ásia e Montanhas Semien, na Etiópia) adaptaram-se a esses ambientes extremos e as baixas concentrações de oxigênio. Visto que locais de alta altitude são caracterizados por duras condições climáticas, tais como: baixas concentrações de oxigênio, altos índices de radiação ultravioleta (UV), períodos de frio extremo, grande amplitude térmica e clima árido.

Desse modo, pode-se considerar que, estas populações nativas, estejam idealmente situadas num contexto adequado para os estudos das adaptações genéticas, dada a onipresença do estresse ambiental ao qual estão submetidas.

A ocupação humana permanente no Altiplano Andino ainda é tema de controvérsias. Algumas evidências arqueológicas remetem a um período que pode variar de 12.000 a 7.000 anos atrás (Rademaker *et al*., 2014; Chala-Aldana *et al*., 2017). É importante ressaltar que estas populações não apenas alcançaram o Altiplano Andino, mas construíram civilizações e permanecem até hoje em tais ambientes, envolvendo um complexo processo que inclui tanto adaptações genéticas quanto culturais (Lindo *et al*., 2018).

Diversas características têm sido associadas ao fenótipo de adaptação à altitude nos Andes, dentre elas, observam-se: o aumento da capacidade pulmonar, a relativa tolerância à hipóxia, e moderado aumento nos glóbulos vermelhos (Frisancho*et al*., 2013). Entretanto, a elucidação completa da natureza dos processos biológicos relacionados à adaptação individual e populacional para a aquisição de tais fenótipos ainda permanece desconhecido.

Os padrões de diversidade genética observados nas populações humanas constituem uma base importante para muitas áreas da pesquisa em genética humana. Mais notavelmente, eles (os padrões) fornecem uma fonte inestimável de dados para inferências sobre nossa história demográfica e evolutiva. Além disso, tais variações genéticas auxiliam também na busca de genes e suas variantes que conferem maior suscetibilidade (ou proteção) às doenças sejam elas crônicas, agudas ou infecciosas.

Tendo em vista o que foi apresentado acima, nesta tese é proposto a investigação de possíveis variantes genéticas envolvidas na adaptação à vida em alta atitude. Para tal finalidade, investigamos diversas populações nativas que vivem desde épocas Pré-Colombianas até os dias atuais nas terras altas dos Andes ou em terras baixas da América do Sul (Mesoamericanos e Amazônicos), como um contraponto. Além disso, foi avaliada a relação entre as variantes genéticas identificadas e as variáveis climáticas que caracterizam os climas aos quais estas populações estão (ou estiveram) submetidas, no intuito de compreender os mecanismos genéticos envolvidos na adaptação à vida em altitude presente nas populações nativas Americanas, ainda insuficientemente compreendidos.

1.2 As populações Nativas Americanas

Através de uma abordagem multidisciplinar integrando dados de arqueologia, genética e linguística, o cenário para a origem humana moderna remonta há cerca de 150.00-200.000 anos antes do presente (Scheinfeldt, *et al*., 2010), na África. Alguns achados arqueológicos, no entanto, apontem para uma origem africana ainda m mais antiga, a cerca de 315.000 anos antes do presente, no Marrocos, sugerindo uma região distinta da África Oriental (Hublin *et al*., 2017). Lamentavelmente, Hublin e colaboradores não conseguiram extrair DNA dos fósseis de Djebel Irhoud. Uma análise genômica poderia ter elucidado e estabelecido se esses fósseis reescrevem verdadeiramente a história da nossa espécie.

Acredita-se que, entre 50 e 100 mil anos atrás, uma pequena fração da população original (hominínios) deixou a África em direção à Europa, Ásia, Oceania, e finalmente América, através da Beríngia, uma grande massa de terra com cerca de 1 milhão de Km² (Santos *et al*., 2007)que ligava a Sibéria com o Alasca durante o Último Máximo Glacial (LGM, da sigla em inglês).

Desde a migração do *Homo sapiens* para fora da África, os humanos modernos colonizaram uma ampla gama de ambientes, desde regiões árticas à tropicais, das terras altas às terras baixas, e até mesmo ambientes considerados tóxicos, em um período relativamente curto de tempo (Schlebusch *et al*., 2015; Fan *et al*., 2016; Henry, 2019). Sendo assim, adaptações culturais, comportamentais e biológicas têm sido necessárias para o sucesso desta trajetória.

As glaciações do Pleistoceno (período que compreende 2.5 milhões a 11.7 mil anos antes do presente) - particularmente o LGM - desempenharam um papel importante na formação dos padrões atuais da diversidade genética humana observados ao redor do mundo. A chegada dos humanos na América é tema de intenso debate na comunidade científica. González-José e colaboradores (2008) propuseram um modelo de povoamento do continente americano integrando resultados de diferentes áreas do conhecimento, tais como, a genética, a morfologia e a linguística. Nesse modelo, dentre outras coisas, o papel da Beríngia é reforçado, já que populações vindas da Ásia permaneceram no local por um período suficiente para o surgimento e fixação de variantes genéticas exclusivas das populações americanas.

As variantes genéticas exclusivas Nativas Americanas podem ser detectadas tanto no genoma nuclear (Schroeder *et al*., 2007; Wang *et al*., 2007), quanto no mitocondrial (Fagundes *et al*., 2008) e no cromossomo Y (Pena *et al*., 1995; Bortolini *et al*., 2003). Estudos recentes analisando genomas de restos humanos oriundos do Alasca do período Pleistoceno Superior, confirmaram partes importantes dos achados anteriores (Moreno-Mayar *et al*., 2018), em especial daqueles propostos por González-José *et al.,*(2008). Durante a permanência dos humanos na Beríngia (Fagundes *et al*., 2008, estimaram um período de cerca de 5.000–8.000 de ocupação), o clima severo e a formação de barreiras glaciais podem ter influenciado o isolamento de populações por longos períodos, e, até mesmo a dificuldades na dispersão populacional pela região. Acredita-se que o aumento populacional, somado à diminuição dos recursos disponíveis devido a "submersão" da

Beríngia, resultou na migração dos primeiros povos para a América - há cerca de 21.000 - 15.000 mil anos antes do presente (González-José *et al*., 2008; Reich *et al.*, 2012; Bortolini *et al*., 2014; Skoglund *et al*., 2015; Posth *et al*., 2018; Moreno-Mayar *et al*., 2018; Pinotti *et al*., 2019).

Após o colapso da Beríngia, no fim do LGM houve fluxo gênico bidirecional, bem menos intenso, mas constante entre os habitantes do extremo oeste do Alasca com aqueles do extremo leste da Sibéria (González-José *et al*., 2008), o que foi posteriormente corroborado com um estudo independente avaliando STRs (*Short Tandem Repeats*) (Ray *et al*., 2009). Esses eventos posteriores ao desaparecimento da Beríngia (surgimento do estreito de Bering) impactaram de forma mais importante às populações nativas do círculo polar ártico e da América do Norte (González-José *et al*., 2008).

Os humanos que colonizaram o continente americano se depararam com uma ampla variedade de ambientes, o qual resultou em adaptações genéticas únicas e também grande diferenciação em termos de suas culturas e sociedades. Basta referir que algumas populações caçadoras coletoras, em períodos Pré-Colombianos, domesticaram plantas e animais, tornando-se agriculturalistas e sedentárias, o qual sustentou a emergência de civilizações que envolviam estados organizados com cidades maiores do que aquelas de onde vieram os colonizadores europeus.

Tendo em vista a enorme amplitude dos ecossistemas observados no continente americano - desde as baixas latitudes do Ártico até elevadas altitudes dos Andes - é esperado que parte da diversidade genética encontrada nas populações nativas seja o resultado de processos não casuísticos, tais como a seleção natural darwiniana (Acuña-Alonzo *et al*., 2010). Contudo, para se ter um panorama completo da história evolutiva das populações ameríndias, deve-se também levar em consideração fatores microevolutivos casuísticos, como a deriva genética e suas derivações (por exemplo, o efeito do fundador), as quais afetam a flutuação aleatória das frequências dos alelos pela introdução ou eliminação casual dos mesmos, independente de seu valor adaptativo. Tal fenômeno se mostra extremamente importante nas populações ameríndias, especialmente aquelas com hábitos de caça e coleta, devido ao baixo tamanho e isolamento populacional por motivos de natureza geográfica e/ou socioculturais.

Apesar disso, estudos têm mostrado que sinais de pressões seletivas ocorreram dentro do continente Americano (Hünemeier *et al*., 2012; Amorim *et al*., 2017). Por

exemplo, Hünemeier e colaboradores (2012) avaliaram a correlação entre os modos de subsistência (caçadores-coletores *versus* agriculturalistas) e as frequências do SNP (polimorfismo de nucleotídeo único) Arg230Cys do gene de metabolismo do colesterol *ABCA1* (*Atp-Binding Cassette, Subfamily A – Member 1*) em populações nativas da América do Sul e Mesoamérica, associando as mesmas à registros arqueológicos de pólen de milho domesticado. Os dados sugeriram que a domesticação de milho foi a força motriz no aumento das frequências do alelo nativo exclusivo 230Cys na Mesoamérica, constituindo um exemplo evidente de coevolução gene-cultura envolvendo um alelo americano autóctone (Hünemeier *et al*., 2012).

Amorim e colaboradores (2017), por sua vez, avaliaram variantes alélicas de genes *FADS* (*Fatty Acid Desaturases*) que já tinham sido descritas por apresentar sinais de seleção positiva em populações Inuit da Groelândia (Fumagalli *et al*., 2015). Neste novo estudo, foram avaliados 349 indivíduos Nativos Americanos: da América do Norte, Mesoamérica e América do Sul. Interessantemente, sinais de seleção natural nos genes *FADS* em populações do Ártico e das Américas sugerem uma assinatura de seleção natural compartilhada entre as populações, indicando que tal adaptação ocorreu na população ancestral comum antes da entrada no Novo Mundo (possivelmente na Beríngia), e, se espalhou por todos os indivíduos do continente americano. Acredita-se que a seleção tenha ocorrido, possivelmente, devido à adaptação genética ao clima frio da região e a uma dieta rica em proteínas e ácidos graxos (Amorim *et al*., 2017).

Por fim, exemplos de seleção natural associados à vida em altitude nos indivíduos do altiplano andino já foram reportados. Interessantemente, diferentes conjuntos de genes apresentam sinal de seleção natural tanto nos Andes, quanto nas outras populações de altitude do mundo, indicando que sob pressões seletivas semelhantes, diferentes soluções genéticas emergiram (Bigham *et al*., 2016; Childebayeva *et al*., 2019). Neste contexto, nos próximos tópicos são apresentados os principais achados científicos sobre as adaptações genéticas às altas altitudes no Platô do Tibete (Ásia), Montanhas Semien da Etiópia (África) e especialmente no Altiplano Andino (América do Sul).

1.3 Adaptação genética à altitude

1.3.1 O que Andinos, Tibetanos e Etíopes têm em comum

Regiões de altas altitudes são locais acima de 2.500 metros acima do nível do mar que desafiam a sobrevivência e a reprodução humana. A principal característica dessas regiões é o ar rarefeito, ou seja, baixa quantidade de oxigênio (O_2) disponível (hipóxia). A nível do mar, a pressão barométrica é de 1 (uma) atmosfera (1 atm), o equivalente a 760 mmHg (milímetro de Mercúrio), e 100% da quantidade de O_2 disponível - que representa 21% dos gases totais da atmosfera. Em contrapartida, a 2.500 m a pressão barométrica é de 570 mmHg e há apenas 75% de O_2 disponível em comparação com o nível do mar (http://www.altitude.org/home.php). Surpreendente, aos 3.821 m de altitude está localizado o lago Titicaca, região reconhecidamente povoada pelos povos Uro, Aymara e Quechua há milhares de anos. Nesta altitude, somente 64% de O_2 disponíveis são observados, e a pressão barométrica atinge somente cerca de 486 mmHg.

Outra característica dos ambientes elevados é a baixa temperatura. A temperatura cai, em média, 1°C (grau Celsius) a cada 150 m. Além disso, outro fator importante é a oscilação de temperatura observada nas regiões de altitude: a amplitude térmica média chega a 20ºC em 24 horas nos Andes (dados próprios, compilados para esta tese). Pode-se acrescentara este cenário, dentre inúmeras outras variáveis climáticas, a baixa umidade do ar e o aumento da radiação ultravioleta (UV), sendo possível ter em mente um cenário não compatível com a vida. No entanto, isso não é observado, pois diversas espécies de plantas, fungos, bactérias e animais estão, há milênios, adaptados a vida em altitude. Desse modo, tais espécies, incluindo o *Homo sapiens*, tornam-se um foco extraordinário para estudos evolutivos.

Seres humanos habitam há anos pelo menos três regiões geograficamente distintas, mas que compartilham o fato de serem consideradas de altas altitudes: O Altiplano Andino (altitude média de 3.750m), continuamente habitado desde o final do Pleistoceno, aproximadamente 12.000 anos antes do presente (Rademaker *et al*., 2014; Chala-Aldana *et al*., 2017), o Platô do Tibete (altitude média de 4.000m) entre a China e a Índia, colonizado no Pleistoceno superior há cerca de 30.000 anos, com migrações posteriores e estabelecimento de aldeias permanentes no planalto há pelo menos 6.500 anos antes do presente (Aldenderfer *et al*., 2011), e ocupações nas terras altas da Etiópia (altitude média de 2.500 m), Montanhas Semien, que teriam ocorrido de 5.000 até, possivelmente, 70.000 anos antes do presente (Alkorta-Aranburu *et al*., 2012). Apesar de algumas incertezas na datação de ocupação permanente destas localidades, é certo que os humanos habitam essas

regiões de climas hostis a milhares de anos. Atualmente, estima-se que cerca de 466 milhões de pessoas vivem permanentemente acima dos 2.500 m de altitude, conforme indica o *Center for International Earth Science Information Network* (http://www.ciesin.org/).

Todos os organismos, desde as bactérias até os seres humanos, possuem mecanismos para manter a homeostase do O_2, que são essenciais para a manutenção da vida na forma que conhecemos. A baixa concentração do O_2 (ou hipóxia) pode resultar na incapacidade de gerar adenosina trifosfato (ATP) suficiente para manter as funções celulares essenciais, enquanto a hiperóxia resulta na geração de intermediários reativos de oxigênio e danos potencialmente letais às membranas celulares e ao DNA. Assim, as concentrações celulares de oxigênio devem ser rigidamente reguladas dentro de uma faixa fisiológica estreita. Desta forma, a contribuição de uma multiplicidade de fatores regulatórios é necessária para estabelecer e modular esses sistemas fisiológicos.

Os mecanismos fisiológicos adaptativos desencadeados pela exposição à altitude têm o objetivo de aumentar a disponibilidade de O_2, aumentando a concentração do mesmo nos tecidos (**Figura 1**). Isto é conseguido através de modificações na: (I) ventilação pulmonar, (II) volume pulmonar e capacidade de difusão pulmonar, (III) transporte de oxigênio no sangue, (IV) difusão de oxigênio do sangue para os tecidos e (V) utilização de O_2 ao nível do tecido (Frisancho, 2013).

(I) O processo de aclimatação fisiológica transiente (*short-time*), aquele que ocorre quando indivíduos de terras baixas migram para altitudes elevadas, inicia-se dentro de 72 horas após a subida, com um aumento progressivo da ventilação pulmonar através da estimulação dos quimiorreceptores arteriais, que chega a ser 100% maior em comparação ao nível do mar. (II) Após a exposição inicial à hipóxia devido à altitude, o volume de ar exalado após uma inspiração vigorosa [conhecida como capacidade vital forçada (FVC) do volume pulmonar] dos nativos das terras baixas é reduzida. No entanto, após cerca de um mês de residência em altas altitudes, esses indivíduos atingem valores comparáveis aos que tinham em baixas altitudes. (III) Em seguida, ocorre o aumento do número de hemácias - em resposta às quantidades insuficientes de oxigênio - a medula óssea é estimulada por um fator eritropoiético para aumentar a produção de hemácias. Por essa razão, em altas altitudes, tanto os nativos das terras baixas quanto os das terras altas têm contagens de hemácias e hemoglobina normalmente superiores às encontradas em indivíduos que se

encontram à nível do mar. (IV) Para que o oxigênio seja utilizado, ele deve atingir as mitocôndrias através de um processo de difusão física. Como o oxigênio é consumido à medida que passa por sucessivas camadas de tecido, a concentração de oxigênio diminui rapidamente, e quanto maior a distância que o oxigênio tem de percorrer, mais rápida é a queda de O_2. Como em altas altitudes a concentração de O_2 já é baixa, o organismo responde encurtando a distância que este deve percorrer, através do surgimento de novos capilares, aumentando a perfusão sanguínea e, assim, facilitando a difusão do O_2 nos tecidos (Frisancho, 2013). (V) O último passo no processo de adaptação à hipóxia envolve variações na taxa de utilização de O_2 e geração de energia no nível celular. Com base em estudos in vivo, foi proposto que em altas altitudes, a glicólise anaeróbica prossegue pela via das pentoses-fosfato, em vez da via de Embden-Meyerhof. A vantagem da via das pentoses parece estar relacionada ao fato de que nenhuma adenosina trifosfato adicional é necessária para gerar gliceraldeído-3-fosfato e assim o organismo economiza energia e produz mais energia química com o mesmo consumo de oxigênio (Frisancho, 2013).

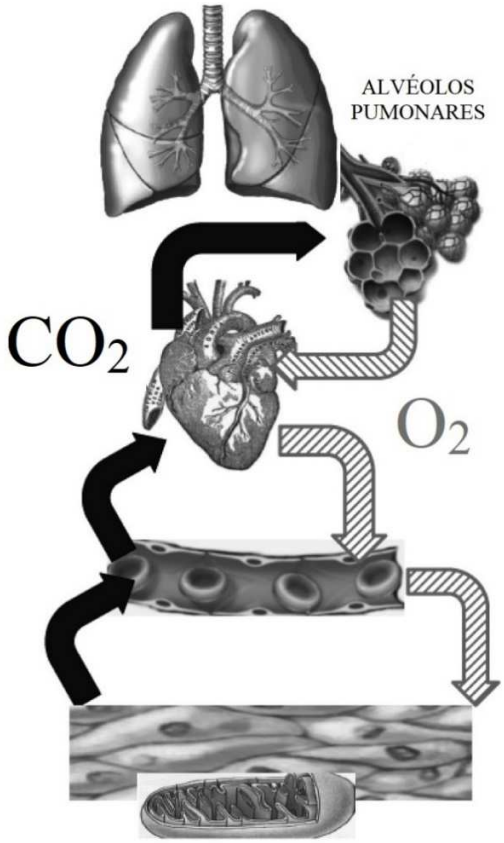


Figura 1. Vias do transporte de O_2 desde a absorção pelos alvéolos pulmonares, até sua utilização para geração de energia na mitocôndria. Adaptado e traduzido de Frisancho, 2013.

Todo o complexo processo fisiológico e metabólico que foi descrito acima é relativamente bem conhecido. Tais respostas funcionais plásticas ocorrem quando indivíduos de baixa altitude visitam ou residem por um curto período em terras altas (Moore, 2017). Porém, sabe-se que esses tipos de adaptações biológicas reversíveis e de curto prazo podem ser relativamente diferentes daquelas que envolvem residentes nativos de longa data (*long-time*). Sob a profundidade temporal de um contexto evolutivo, várias características morfológicas, metabólicas e fisiológicas duradouras são necessárias para uma vida de sucesso em altas altitudes.

Desta forma, espera-se que a seleção natural deixe "pegadas" no genoma dos indivíduos que habitam essas regiões caracterizadas pelas condições climáticas extremas. Esses grupos humanos superaram a baixa concentração de O_2 no ambiente de altas altitudes e, um estudo mais aprofundado dos mecanismos que governam as mudanças adaptativas responsáveis pela adaptação à grandes altitudes contribuirá para o entendimento das bases moleculares da mudança evolutiva e auxiliará na anotação funcional do genoma humano.

Recentemente, vários estudos têm revelado as bases genéticas dessas mudanças nas populações que vivem em altitude, acima descritas. O mecanismo exato pelo qual se deu tal adaptação ainda não se encontra completamente elucidado, no entanto, diversos estudos tentam entender o papel de várias variantes genéticas associadas à adaptação à altitude (Yi *et al*., 2010; Scheinfeldt *et al*., 2012; Huerta-Sánchez *et al*., 2014; Simonson *et al*., 2015; Valverde et al., 2015; Fehren-Schmitz and Georges, 2016; Crawford *et al*., 2017; Moore, 2017).

Uma via genética que recorrentemente tem sido implicada na adaptação humana à alta altitude é a via do fator de indução por hipóxia (HIF), que é o principal regulador da resposta transcricional à hipóxia em metazoários desde o início do desenvolvimento embrionário e ao longo da vida adulta (Semenza, 1999). Na **tabela 1**, observamos os prncipais membros da via de HIF.

Tabela 1. Membros da via do fator de indução por hipóxia (HIF)

α	**β**
HIF-1α	HIF-1β (ARNT)

HIF-2α	ARNT2
HIF-3α	ARNT3

Disponível em Semenza, 1999.

HIF-1 foi purificado e caracterizado através de um ensaio de cromatografia por Wang e Semenza em 1995. HIF-1 é um heterodímero que consiste em uma subunidade β (HIF-1β, também conhecida como ARNT, (*Aryl Hydrocarbon Receptor Nuclear Translocator*) e três subunidades α (HIF-1α, HIF-2α e HIF-3α). Através de análises de sequências nucleotídicas de clones cDNA, descreveu-se que HIF-1 contém um domínio N-terminal HLH *(Helix-Loop-Helix*) seguido pelo domínio PAS (*Per Arnt Sim*). O domínio HLH medeia a dimerização e ligação com o DNA num grande número de fatores de transcrição, enquanto que o domínio PAS desempenha um importante papel na formação de heterodímeros. As subunidades α (HIF-1α, HIF-2α e HIF-3α) contém, na porção C-terminal, domínios de ativação transcricional e degradação dependentes de O_2 (Semenza e Wang, 1995). Em resposta à concentração de O_2 disponível, o principal meio pelo qual a atividade de HIF é controlada é através da hidroxilação sítio-específico da subunidade α (Semenza, 1999).

Sob condições de normóxia, uma família de três enzimas, PHD1 (*Hypoxia-InducibleFactor Prolyl Hydroxylase-1*), também conhecida como EGLN2. PHD2, também conhecida como EGLN1, e PHD3, também conhecida como EGLN3 - hidroxilam a subunidade α em seu domínio de degradação dependente de O_2. Essa hidroxilação permite a associação da proteína von Hippel-Lindau (VHL) que faz parte de um complexo E3 ubiquitina ligase que, em seguida, encaminha HIF-α para a degradação via ubiquitina-proteossomo. Em um cenário de hipóxia, essa modificação dependente de O_2 é interrompida, levando à dimerização de HIF-α com HIF-β, e esntão este complexo se liga aos elementos responsivos a hipóxias (HREs) que vão atuar como controladores dos genes alvos (**Figura 2**) (Bigham e Lee, 2014).

Fatores relacionados com HIF-1α, que dimerizam com HIF-β, já foram descritos: HIF-2α, também chamado de EPAS1 (*endothelial PAS domain protein 1*). HIF-2α é semelhante ao HIF-1α no que diz respeito não só a estrutura, mas também nos principais domínios funcionais e na regulação da atividade de hipóxia. Como resultado, observa-se a ativação da transcrição de mais de cem (100) genes possuindo elementos responsivos à

hipóxia em sua região promotora. Dos quais podemos citar: genes envolvidos na angiogênese (*VEGF*, *PDGF*, *FLT*), eritropoiese (*EPO*), morte celular (*TGFA*, *TNF*, *MMP*), inflamação (*IL6*), vasorreatividade (*EDN*, *NOS*, *PRKAA1*), metabolismo de glicose (*SLC2A1*, *IGFBP*, *PRKAA1*), bem como outros processos que afetam as várias etapas do sistema de transporte de O_2 (Greer *et al*., 2012; Moore, 2017).

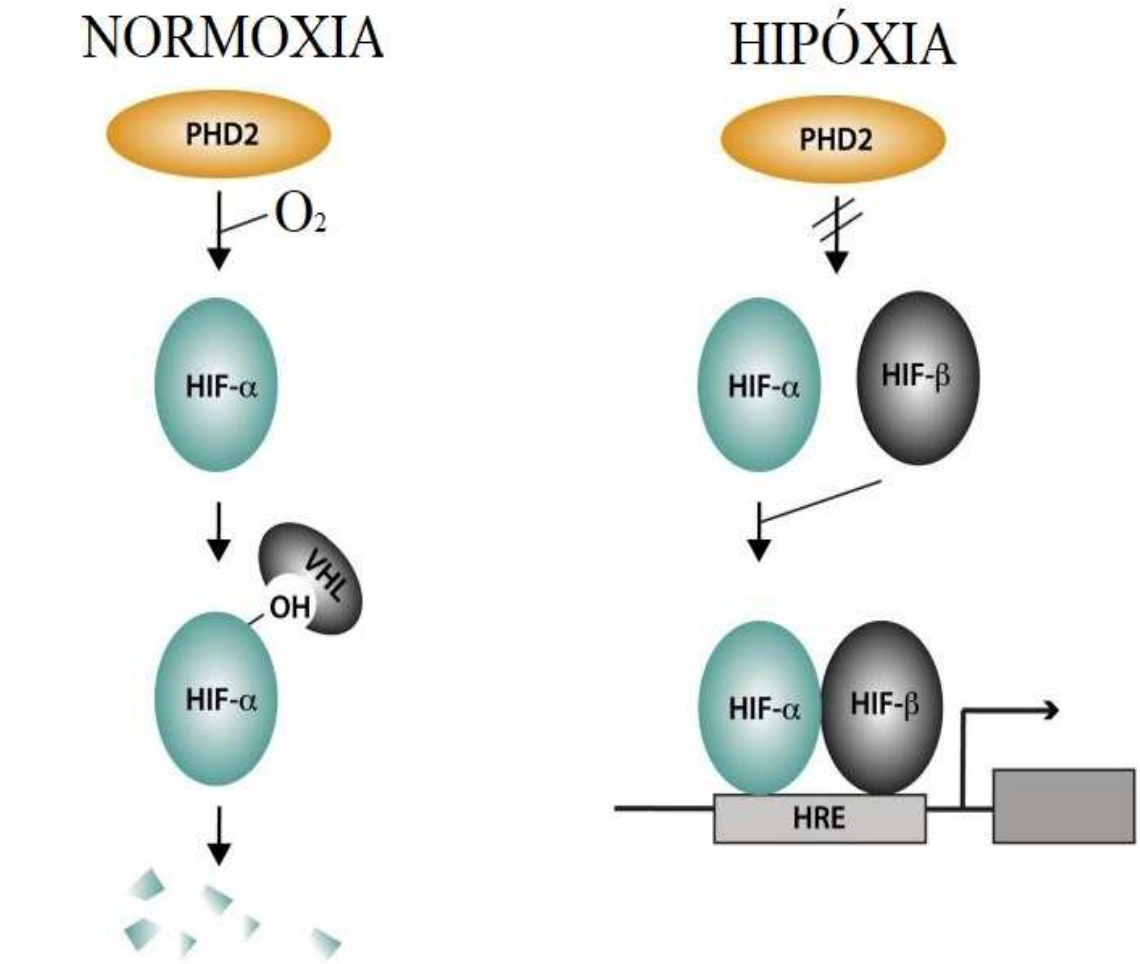


Figura 2: Via de HIF. À esquerda, em condições normais de O_2, PHD2 (*Hypoxia-Inducible Factor Prolyl Hydroxylase 2*; conhecido como EGLN1) hidroxila constitutivamente HIF-α, visando sua degradação de maneira dependente de VHL (*von Hippel-Lindau*). À direita, em hipóxia, a hidroxilação é interrompida, ocasionando a estabilização de HIF-α, dimerização com HIF-β e ligação a elementos de resposta à hipóxia (HREs) que controlam genes alvo. Adaptado e traduzido de Bigham e Lee, 2014.

HIF-1α e HIF-2α são ativados conjuntamente, mas desempenham distintas respostas celulares. HIF-1α regula positivamente os genes que codificam as enzimas glicolíticas e, inibe a fosforilação oxidativa e a biogênese mitocondrial (Semenza, 1999). HIF-2α por sua vez, desempenha um papel importante na regulação do gene *EPO* (*Erythropoietin*) em células intersticiais especializadas no rim adulto (Lee e Percy 2011). A eritropoietina é o regulador crítico envolvido na produção de glóbulos vermelhos (Scortegagna *et al*., 2005).

Como é esperado dos fatores que orquestram a resposta à hipóxia, HIF-1α e HIF-2α geralmente atuam na mesma direção. Ambos têm a capacidade de ativar o gene *VEGFA*(*Vascular Endothelial Growth Factor A*), o qual codifica uma proteína-chave que induz a angiogênese e já foi implicado na patogênese da hipertensão pulmonar. Similarmente, tanto HIF-1α quanto HIF-2α foram implicados na patogênese da hipertensão pulmonar (Keith *et al*., 2011; Shimoda e Laurie, 2014). No entanto, também está claro que HIF-1α e HIF-2α podem ter atividades antagônicas. Por exemplo, dependendo do contexto celular ao qualestão expostos, HIF-1α pode promover a parada do ciclo celular, enquanto HIF-2α mantém a progressão ao longo do ciclo celular (Keith *et al*., 2012).

Inúmeros exemplos de evolução convergente na via da HIF têm sido documentados em populações humanas de altas altitudes (Scheinfeldt e Tishkoff, 2010; Simonson *et al*., 2012). Nas populações do altiplano andino, *PRKAA1*, *EPAS1*, *VEGF*, *NOS2*, *TGFA*, *CXCR4* e *EGLN1* já foram descritos com sinal de seleção natural (Bigham *et al*., 2009; Bigham *et al*., 2010; Foll *et al*., 2014; Eichstaedt *et al*., 2017), enquanto nos tibetanos, *EPAS1* e *EGLN1* foram extensivamente identificados (Beall *et al*., 2010; Yi *et al*., 2010; Peng *et al*., 2011; Lorenzo *et al*., 2014; Simonson *et al.*, 2015; Hu *et al*., 2017; Lanikova *et al*., 2017;Tashi *et al*.,2017; Jeong *et al*., 2018). Estudos focados em populações das Montanhas de Semien, na Etiópia, sugerem o envolvimento de uma combinação de genes *HIF* diferente dos encontrados em Andinos e Tibetanos: *THRB*, *ARNT2*, *BHLHE41* (gene induzido por HIF-1α) e *RORA*, que codifica uma proteína que induz a ativação transcricional de HIF-1α (Alkorta-Aranburu *et al*., 2012; Scheinfeldt *et al*., 2012; Huerta-Sánchez *et al*., 2014).

Na **Figura 3** podemos observar os órgãos envolvidos na adaptação à altas altitudes, juntamente com genes selecionados da via de HIF e genes-alvo de HIF.

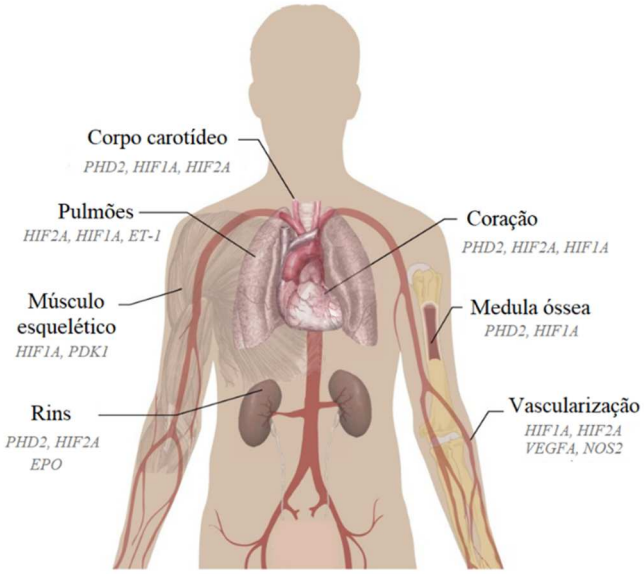


Figura 3. Órgãos envolvidos na adaptação à altas altitudes, juntamente com genes selecionados da via de HIF e genes-alvo de HIF que podem desempenhar um papel relevante na adaptação à altitude. Adaptado e traduzido de Bigham e Lee, 2014.

Na ilustração abaixo (**Figura 4**), extraída do trabalho de Abigail W. Bigham (2016), pode-se observar destacadas em azul as três regiões globais de alta altitude. Os genes candidatos da via de HIF para cada população são listados em cada uma das populações.

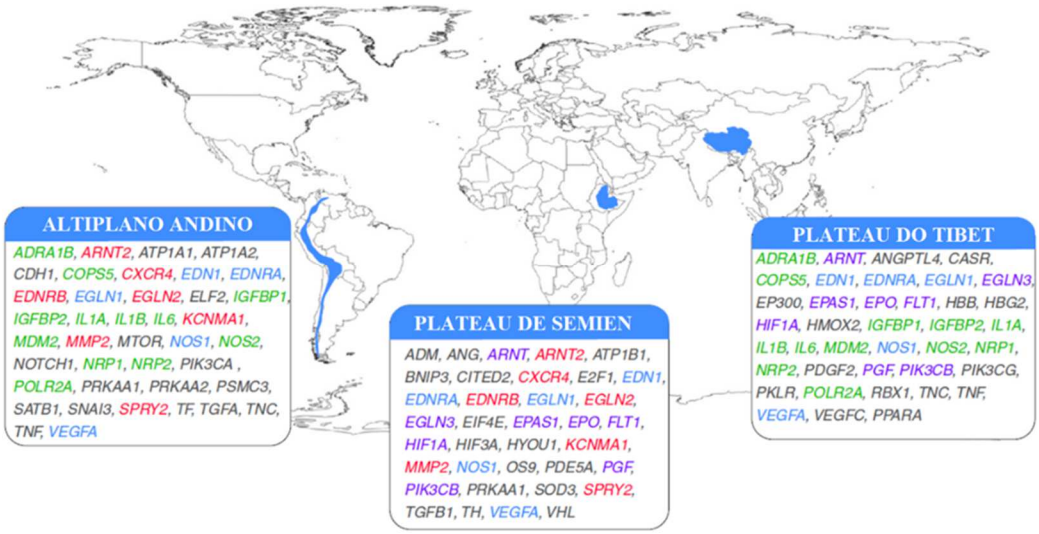


Figura 4. Genes da via de HIF com papel na adaptação humana às grandes altitudes. Os genes listados em preto são únicos para cada população em particular, os azuis são compartilhados entre as três populações, os

verdes são compartilhados entre Tibetanos e Andinos, os vermelhos são compartilhados entre Andinos e Etíopes e os roxos são compartilhados entre Tibetanos e Etíopes. Adaptado e traduzido de Bigham, 2016.

Como se pode observar, apesar de diversos genes compartilharem sinais de seleção natural, cada população de altitude possui características únicas no que se refere às adaptações genéticas, o que é esperado na evolução convergente. Sendo assim, as populações Etíopes, Andinas e Tibetanas que vivem em altitudes elevadas se adaptaram à hipóxia de maneira um tanto quanto diferente, afetando diferentes genes de uma mesma via (ou de rotas relacionadas), ou ainda os mesmos genes, mas com variantes exclusivas e/ou conjunto de alelos comuns, porém com marcantes diferenças em suas distribuições.Nos tópicos seguintes, são detalhados os mecanismos genéticos específicos de cada uma das populações de altas altitudes.

1.3.2 Adaptação genética à altitude: Os Tibetanos

Os indivíduos que vivem no platô do Tibete são - sem sombra de dúvida - os mais extensivamente estudados no que diz respeito às adaptações genéticas para vida em altas altitudes. Diversos estudos demonstraram variantes adaptativas nas regiões dos genes *EGLN1* e EPAS1 (ou HIF-2α), este último associado a níveis mais baixos de hemoglobina (Beall *et al*., 2010; Simonson *et al*., 2010; Yi *et al*., 2010).

Através de experimentos *in vitro*, Song e colaboradores (2014) descreveram que duas substituições não sinônimas na região codificadora de *EGLN1* [éxon 1: rs12097901 (C127S) e rs186996510 (D4E)] resultam numa menor interação com proteína p23, afetando negativamente a regulação do *EGLN1* na via do HIF. É proposto que estas variantes tibetanas ocorrendo juntas em forma de haplótipo (D4E/C127S) gerem um alelo de perda de função (hipomórfico), levando a uma potencialização na sinalização de HIF, assim facilitando a adaptação à altas altitudes (Song *et al*., 2014).

Um segundo gene candidato da via do HIF com evidências de seleção positiva nos Tibetanos em diferentes estudos é o *EPAS1* (Beall *et al*., 2010; Simonson *et al*., 2012; Y*i et al*., 2010; 2014; Peng *et al*., 2017; Yang*et al*., 2017; Hu *et al*., 2017; Jeong *et al*., 2018). Este gene codifica o fator de transcrição HIF-2α, que estimula a produção de glóbulos vermelhos (**veja tópico 1.3.1**). Além disso, variantes de *EPAS1* encontradas em Tibetanos foram associadas a baixas concentrações de hemoglobina, sugerindo um papel importante na proteção à policitemia (Beall *et al*., 2010; Yi *et al*., 2010).

Adicionalmente, estudos recentes em camundongos heterozigotos *knockout* para o gene *EPAS1*$^{+/-}$ exibem respostas fisiológicas enfraquecidas à hipóxia crônica, com redução de aproximadamente 50% no nível de transcritos íntegros. Além disso, essas variantes de *EPAS1* não estão associadas apenas ao nível relativamente baixo de hemoglobina como um protetor de policitemia (superprodução de eritrócitos), mas também está associada a uma vasoconstrição pulmonar diminuída nesses indivíduos, possibilitando a manutenção de uma capacidade de ventilação pulmonar adequada. Tomados em conjunto esses dados mostram que a regulação negativa de *EPAS1* contribui para a base molecular da adaptação pelo menos dos Tibetanos à hipóxia em altas altitudes (Peng *et al*., 2017).

Huerta-Sánchez e colaboradores (2014), sequenciaram o gene *EPAS1* e regiões no entorno do mesmo, em 40 indivíduos nativos do Tibete e 40 indivíduos chineses da etnia Han, que habitam terras baixas. Através de análises comparativas, observaram que o gene possui uma estrutura haplotípica incomum nos Tibetanos. Expandindo as análises para um conjunto maior de populações ao redor do mundo, os autores descreveram que o haplótipo selecionado é encontrado apenas no hominínio de Denisova e em Tibetanos e, em uma frequência muito baixa entre os chineses Han. A extensão do haplótipo, e o fato de não ser encontrado em outras populações mundiais, torna improvável que o compartilhamento de haplótipos entre Tibetanos e o povo de Denisova tenha sido causado por uma simples ancestralidade comum, sugerindo assim uma introgressão adaptativa de uma porção do gene *EPAS1*, a partir de "denisovanianos" para populações de *Homo sapiens* que habitavam a mesma região. Essa descoberta reforça que a hibridização do *Homo sapiens* com hominínios arcaicos teria sido elemento chave como fonte de variabilidade genotípica que, em certas ocasiões, pode ter auxiliado os humanos modernos a se adaptarem a novos ambientes (Huerta-Sánchez *et al*., 2014).

Outro gene interessantemente encontrado com sinal de seleção natural nas populações Tibetanas é o *MTHFR* (*Methylene Tetra Hydro folate Reductase*) cuja proteína codificada - o metilenotetrahidrofolato redutase - consiste em uma enzima que desempenha um papel central no metabolismo do folato. O SNP rs1801133 foi encontrado com sinal de seleção positiva nos tibetanos e, o alelo rs1801133A em maior frequência nas populações tibetanas está diretamente relacionado com o aumento de folato, sugerindo adaptação à alta radiação ultravioleta, visto que a degradação do folato pode ser acelerada pela exposição aos raios UV (Borradale e Kimlin, 2012; Yang *et al*., 2017).

Além dos genes *EGLN1* e *EPAS1* - alvos de seleção diretamente envolvidos na via de detecção de hipóxia, extensivamente identificados com sinal de seleção natural nos tibetanos - outros genes foram descritos. O gene *PPARA* (*Peroxisome Proliferator-Activated Receptor A*), um regulador principal da oxidação de ácidos graxos, foi identificado como um candidato de seleção associado com baixos níveis de hemoglobina nos tibetanos (Simonson *et al*., 2010). Assim, é possível que alterações em genes que regulam o metabolismo, como o *PPARA*, possam compensar as mudanças inerentes às alterações na via do HIF.

Adicionalmente, o gene *HMOX2* (*Heme Oxygenase 2*) envolvido no catabolismo da heme, foi descrito abrigando variantes potencialmente adaptativas em tibetanos. É descrito que nessas populações o alelo derivado (C) de rs4786504 (T>C) confere menores níveis de hemoglobina quando comparados aos portadores do alelo T. Adicionalmente, o alelo derivado (C) é associado a maior expressão de HMOX2 *in vitro*, presumivelmente favorecendo a quebra do heme e auxiliando na manutenção dos baixos níveis de hemoglobina em altas altitudes. Desta forma, os dados sugerem que *HMOX2* contribui para a adaptação a alta altitude em tibetanos, atuando como um modificador na regulação do metabolismo da hemoglobina (Yang *et al*., 2016).

Wang e colaboradores (2011) também descreveram outros genes com sinais de seleção natural em tibetanos, incluindo *ANGPT1* (*Angiopoietin-1*), *ECE1* (*Endothelin Converting Enzyme 1*), e LEPR (*Leptin Receptor*), que são associados a diversas funções biológicas, tais como: o desenvolvimento de vasos sanguíneos, placenta, embriões e gônadas femininas. As funções alteradas desses genes certamente afetariam as gestações e sobrevivência infantil, resultando em fenótipos mal adaptativos nos tibetanos (Wang *et al*., 2011).

Além dos genes previamente descritos, ainda observa-se sinais de seleção natural nos genes: *PTGIS* (*Prostaglandin I2 Synthase*), *VDR* (*Vitamin D3 Receptor*), *KCTD12* (*K+ Channel Tetramerization Domain Containing 12*), *FOXO1* (*Forkhead Box O1*), *RUNX1* (*Runt Related Transcription Factor 1*), *RYR1* (*Ryanodine Receptor 1*) e *TED* (*Tibetan Enriched Deletion*) associados às mais diversas vias biológicas (Wang *et al*.,2011; Lou *et al*., 2015; Hu *et al*., 2017).

Em um recente estudo avaliando mulheres nativas das altas altitudes do Himalaia e Nepal, Jeong e colaboradores (2018) associaram dados genotípicos com dados fenotípicos

que buscavam caracterizar a história reprodutiva dessas mulheres, como o número de filhos nascidos vivos, natimortos, perdas gestacionais, número de nascimento de gêmeos, entre outras. Foi descrito forte sinal de seleção natural nos genes *EPAS1* e *EGLN1,* além de seleção positiva e associação entre os alelos dos genes *CCDC141* (*Coiled-Coil Domain Containing141*), *PAPOLA* (*Alpha*), *VRK1* (*Vaccinia Related Kinase 1*), *C6ORF195* (C*hromosome Poly(A) Polymerase Alpha*), *VRK1* (*Vaccinia Related Kinase 1*), *C6ORF195* (*Chromosome 6 Open Reading Frame 195*), *CTBP2* (*C-Terminal Binding Protein 1*), *TEX36* (*Testis Expressed 36*) e *EDRF1* (*Erythroid Differentiation Regulatory Factor 1*) com fenótipos de sucesso, como aqueles associados à menor mortalidade de filhos, mais gestações e nascidos vivos. Isso demonstra que a seleção natural alterou sistematicamente a frequência dos alelos associados aos resultados reprodutivos em direção ao aumento de *fitness* (Jeong *et al*., 2018).

Podemos observar que os mecanismos de adaptação à vida no Tibete sugerem que as adaptações fisiológicas observadas são controladas através da interação entre múltiplos genes, especialmente, aqueles que constituem a via de HIF.

1.3.3 Adaptação genética à altitude: Os Etíopes

Desafortunadamente, as populações de altitude da Etiópia (Montanhas Semien) ainda são as menos estudadas no que tange a temática de adaptação genética à alta altitude, não obstante o sucesso estraordinário que atletas nativos dessa região tem em competições que envolvem corridas de longa distância, de resistência. No entanto, os poucos estudos disponíveis avaliando essas populações, não mostram um cenário muito diferente das vias encontradas com sinal de seleção positiva em tibetanos e andinos, mas evidentemente, apresentam suas particularidades. Por exemplo, a concentração de hemoglobina elevada em comparação aos valores normais encontrados ao nível do mar foi,durante muito tempo considerada uma característica de adaptação à hipóxia em altas altitudes (Winslow e Monge, 1987). No entanto, estudos em andinos, tibetanos e etíopes demonstram que esta não é uma resposta imprescindível à hipóxia ambiental.

As principais singularidades das populações etíopes encontram-se nos níveis de hemoglobina e saturação de O_2 (SaO_2) dessas populações, que são praticamente iguais às encontradas nas populações que vivem no nível do mar (**Figuras 5 e 6**). Os nativos da Etiópia mantêm as concentrações de hemoglobina venosa e a SaO_2 arterial dentro dos limites das populações do nível do mar, apesar da inevitável diminuição na concentração

de O_2 do ambiente de altas altitudes (Beall *et al.,* 2002). Os resultados deste estudo sugerem que os nativos etíopes de altas altitudes respondem à hipóxia diferentemente dos andinos ou tibetanos. Mas assemelham-se a amostras do planalto tibetano, exibindo pouca ou nenhuma elevação da concentração de hemoglobina nesta faixa de altitude, porém diferenciam-se dos tibetanos apresentando maior saturação de oxigênio (Beall *et al*., 2002).

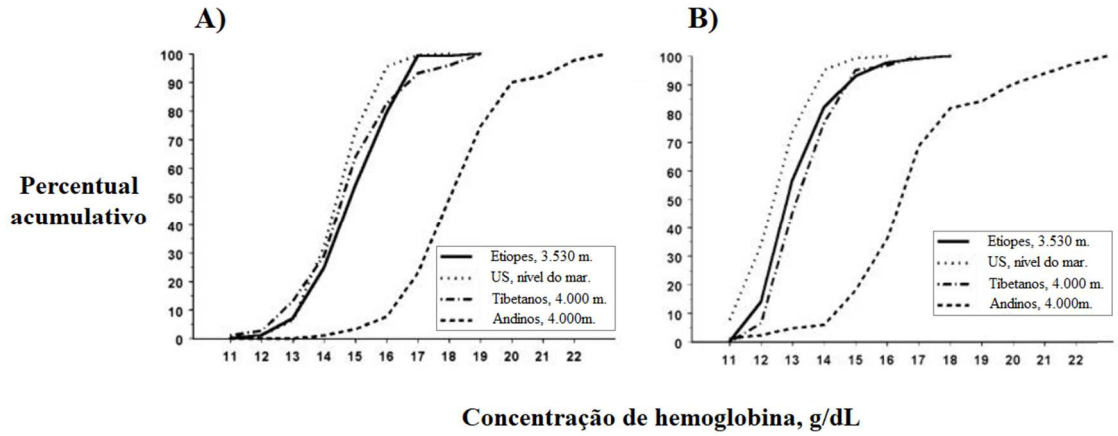


Figura 5. Os gráficos acima mostram a distribuição da concentração de hemoglobina napopulação etíope, norte americanos que vivem à nível do mar, tibetanos e andinos, respectivamente. Em A) homens e B) mulheres. No eixo x observamos a concentração de hemoglobina em g/dl, enquanto que no eixo y encontramos a frequência do percentual cumulativo. Adaptado e traduzido de Beall *et al*., (2002).

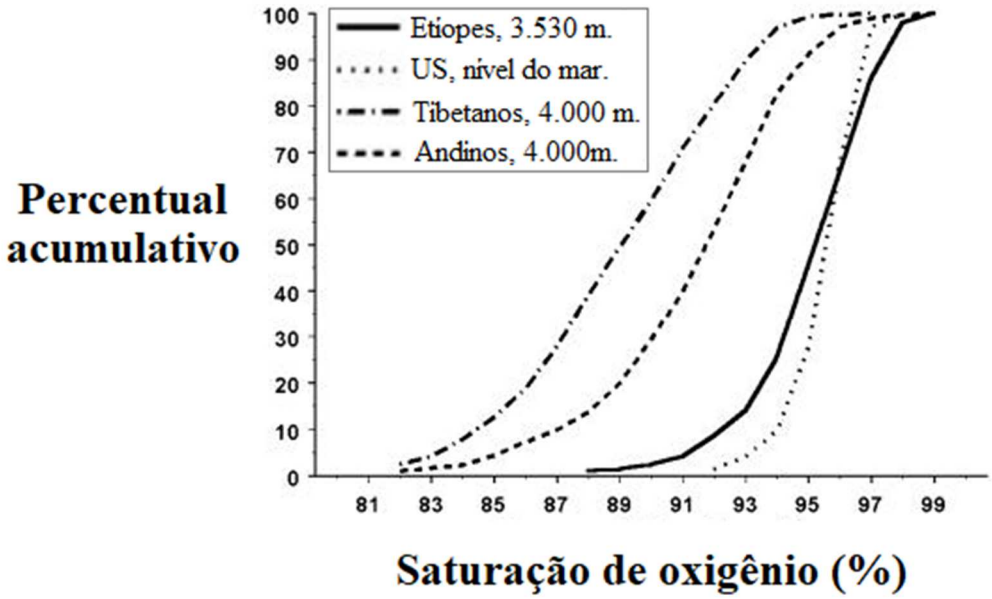


Figura 6. As distribuições de saturação de oxigênio (SaO_2) dos norte-americanos que vivem à nível do mar coincidem com a dos etíopes de altas altitudes e contrastam com os índices reduzidos observadas nos tibetanos e andinos. Adaptado e traduzido de Beall *et al*., (2002).

Na tentativa de compreender melhor esse fenômeno, e elucidar quais os mecanismos biológicos e genéticos que permitem a adaptação bem-sucedida dos etíopes às grandes altitudes, Scheinfeldt e colaboradores (2012) realizaram um estudo de associação com dados genéticos de varredura genômica (GWAS) em populações etíopes de altas altitudes, Amhara (3.202m) e terras baixas da Etiópia (Aari1, 1.407m e Hamer, 1.097m) (Scheinfeldt et al., 2012). As análises sugeriram vias significativamente enriquecidas e com sinal de seleção positiva nos Amhara: *CBARA1* (*Calcium-Binding Atopy-Related Autoantigen 1*), o qual está envolvido na regulação da captação de cálcio pelas mitocôndrias, *VAV3* (*Vav Guanine Nucleotide Exchange Factor 3*) que já foi descrito envolvido na angiogênese *in vivo* - processo que é iniciado pelo HIF-1 - em situações de hipóxia na idade adulta. *ARHGAP15* (*Rho GTPase Activating Protein 15*), o qual pode estar envolvido na sobrevivência após lesão pulmonar aguda e o *RNF216* (*Ring Finger Protein 216*) que codifica uma enzima que inibe as vias de ativação do NF-kappa B envolvidas na indução de HIF-α (Scheinfeldt *et al*., 2012).

Adicionalmente, *ARNT2* também conhecido como HIF-1β, e, *THRB* (*Thyroid Hormone Receptor Beta*) foram encontrados com padrões de variação polimórficas consistentes com sinal de seleção positiva. Estes genes possuem variantes genéticas associadas com os níveis de hemoglobina e participamda cascata de sinalização da via do HIF-1 iniciada sob hipóxia. A proteína THRB é expressa em células hepáticas formando um heterodímero com receptor retinóide X (RXR) necessário para a expressão do HIF. Durante o desenvolvimento fetal, o fígado é a principal fonte de eritropoietina - HIF regula a produção de eritropoietina - que é necessária para a produção de glóbulos vermelhos. O HIF-1α e o ARNT2 formam um heterodímero presente na maioria das células e desempenha um papel geral na resposta à hipóxia (**veja Figura 2**) (Scheinfeldt *et al*., 2012).

Em relação aos níveis de hemoglobina, contrariando o que foi observado por Beall e colaboradores em 2002, foram descritas diferenças significativas entre níveis de hemoglobina de homens Amhara (16,4 g/dl) *versus* Aaril (14,8g/dl) e Hamer (12,4g/dl) (Scheinfeldt *et al*., 2012). Tais diferenças podem ser compreendidas levando-se em conta diversos fatores que podem influenciar os resultados, como as diferentes metodologias, as

quais afetam a sensibilidade da quantificação de hemoglobina entre os dois estudos, o tamanho amostral reduzido no estudo de Scheinfeldt e colaboradores (2012), entre outras possíveis variáveis.

Alkorta-Aranburu e colaboradores (2012), por sua vez, avaliando duas populações das terras altas da Etiópia (Amhara e Oromo) descreveram o sinal de seleção positiva no gene *RORA* (*RAR Related Orphan Receptor A*). Tal achado, revelou um excelente gene candidato para fenótipos de resposta à hipóxia uma vez que pertence à via do HIF-1. Além deste, outros genes candidatos à hipóxia, como *COL6A1 (Collagen Type 6 Alpha-1)*, *SLC30A9 (Solute Carrier Family 30 Member 9)* e *HGF (Hepatocyte Growth Factor)* foram também destacados (Alkorta-Aranburu *et al*., 2012).

Avaliando dados de estudos de associação com dados de GWAS em populações etíopes que vivem em altitude, foram descritos sinais de seleção natural no gene *BHLHE41* (*Basic Helix-Loop-Helix Family Member e41*) (Huerta-Sanchez *et al*., 2014). Este gene é transcricionalmente regulado pelo HIF-1α, o qual posteriormente se liga a HIF-1α a fim de reprimir outros alvos transcricionais induzidos pela hipóxia, incluindo o *VEGF*, provavelmente devido à maior degradação das proteínas HIF-1α e HIF-2α induzido por BHLHE41 (Miyazaki *et al*., 2002).

Em outro estudo, com base na divergência da frequência alélica observada em alguns polimorfismos no gene *HLA-DRA* (*Human Leukocyte Antigen-DR Alpha*) em indivíduos Amhara, foi sugerido um papel de genes de defesa imunológica na adaptação a altas altitudes (Fort et *al.,* 1998). Interessantemente, doenças como a malária e esquistossomose são prevalentes nas comunidades de terras baixas, mas inexistentes nas terras altas. De fato, estudos epidemiológicos na Etiópia mostram a correspondência entre altitude e incidência de malária (Woyessa *et al*., 2013).

1.3.4 Adaptação genética à altitude: Os Andinos

Considerando as populações do Altiplano Andino, várias características têm sido associadas ao fenótipo de adaptação à altitude, dentre elas pode-se citar: o aumento da capacidade pulmonar, tolerância à hipóxia, elevadas concentrações nos níveis de hemoglobina, saturação de oxigênio e ácido nítrico, bem como baixo peso ao nascer (Bigham e Lee, 2014).

Tradicionalmente, o estudo dos fenótipos adaptativos de altas altitude, se concentra principalmente em fatores pulmonares e hematológicos, tais como: índices de hemoglobina, SaO_2 e capacidade pulmonar. No entanto, em novas abordagens o papel dos fatores vasculares vem se destacando como parte importante dos processos adaptativos à altas altitudes. Neste contexto, o óxido nítrico (NO) é um poderoso vasodilatador expresso em vários tipos de células, incluindo as células endoteliais vasculares, e desempenha um papel central na regulação do fluxo sanguíneo e da resistência vascular (Beall *et al*., 2012). O aumento da quantidade de NO na parte interna dos vasos sanguíneos leva ao aumento do fluxo sanguíneo. Desta forma, o gás (NO) se espalha pelo sangue e forma nitritos e nitratos fazendo com que as artérias e capilares se expandem e transportem o sangue rico em O_2 para o restante do corpo mais rapidamente que o normal (Beall *et al*., 2012). Recentemente, um polimorfismo (rs1799983, G > T na posição 894) no gene *NOS3 (Nitric Oxid Synthase 3)*, que codifica a sintase endotelial do óxido nítrico (eNOS), foi descrito com sinal de seleção natural em 109 indivíduos peruanos de terras altas, que teriam vivido a aproximadamente 8.500-5.500 anos antes do presente (Fehren-Schmitz e Georges, 2016).

Outras alterações vasculares observadas nas mulheres andinas incluem aquelas associadas com a gestação e o crescimento fetal. Sugere-se que a hipóxia severa encontrada nos ambientes de altitude retarda o crescimento fetal por reduzir o fluxo sanguíneo da artéria uterina (Moore *et al*., 2001). A restrição do crescimento intrauterino (RCIU) – principalmente baixo peso ao nascer - é um importante marcador de morbidade e mortalidade em recém-nascidos. No entanto, observa-se que bebês nascidos de mulheres andinas são menos afetados pela hipóxia, embora o peso em média desses bebês ainda seja significantemente menor do que os bebês de indivíduos de terras baixas. É fato conhecido que bebês que nascem em altitude tem menor peso ao nascer (cerca de 102 gramas para cada 1.000 m de elevação) (Moore *et al*., 2001). A altitude elevada reduz o peso corporal e comprimento ao nascimento. No entanto, este fenômeno não é observado nos andinos, sugerindo uma influência genética para este fenótipo (Soria *et al*., 2013).

Um gene em particular, relacionado à hipóxia, a subunidade catalítica α-1 da proteína cinase ativada por monofosfato de adenosina (*PRKAA1*: *Protein Kinase AMP-Activated Catalytic Subunit Alpha 1*, também conhecida como AMPKα1), pode desempenhar um papel importante na adaptação genética para altas altitudes, afetando as

respostas fisiológicas à gravidez que são fundamentais para o crescimento fetal, como o diâmetro da artéria uterina (Bigham*et al*., 2014).

Avaliando o panorama geral de estudos com abordagem GWAS e adaptação às grandes altitudes, os andinos foram comparativamente menos investigados, em relação aos tibetanos. Apesar disso, muitos genes candidatos foram claramente identificados, sendo os mais importantes aqueles envolvidos na via de HIF, essa última como já visto na presente tese, desempenha um papel importante na resposta à hipóxia (Bigham *et al*., 2009). O gene *EGLN1* foi reportado, como um gene comum aos andinos e tibetanos, estando sob evolução convergente, mas com ambas as populações exibindo um distinto haplótipo dominante em torno deste gene (Bigham *et al*., 2009).

O grupo de pesquisa liderado por Laurent Excoffier também relatou um processo de evolução convergente na adaptação à altitude em populações andinas e tibetanas (Foll *et al*., 2014). Os autores usaram um método hierárquico Bayesiano capaz de detectar seleção positiva mesmo em complexos cenários demográficos. Eles sugerem que pressões seletivas similares nos Andes e no Himalaia provavelmente ativaram respostas adaptativas a pressão da seleção natural nas mesmas rotas genéticas, especificamente em duas vias que atuam na contenção da toxicidade promovida pela hipóxia. Um dos genes com maior destaque é o *EPAS1* (também conhecido como HIF-2α), localizado no cromossomo humano 2, na região 2p21, sendo constituído por aproximadamente 93Kb (kilobases), compreendendo um total de 16 éxons (http://atlasgeneticsoncology.org/). *EPAS1* também foi reportado por Excoffier e colaboradores como sendo uma região gênica sugestiva de adaptação convergente em andinos e tibetanos. Devido à natureza poligênica da resposta adaptativa às altas altitudes, é muito provável que vários *loci* genéticos estejam envolvidos em uma ou mais redes gênicas, podendo ou não estar interconectadas. Portanto, cada *locus* geralmente possui uma pequena, mas significativa contribuição para o fenótipo complexo envolvido na adaptação a ambientes extremos. Além disso, o estudo reforça a idéia de que tais fenótipos surgirama de maneira concomitante e independente em diferentes continentes, como aqueles encontrados na América (Andes) e na Ásia (Himalaia) (Foll *et al*., 2014).

Interessantemente, nos andinos, o produto do gene *FAM213A* (*Family With Sequence Similarity 213, Member A*), encontrado com sinal de seleção positiva, pode atuar como um potente antioxidante a fim de diminuir o estresse oxidativo e auxiliar na

manutenção da massa óssea. Outro gene detectado com sinal de seleção nos andinos é o *SFTPD* (*Surfactant Protein D*), que codifica uma proteína associada ao surfactante pulmonar envolvida na respiração normal e na defesa imunológica inata do hospedeiro. Esses dados identificam dois novos genes candidatos e vias associadas que podem estar envolvidos na adaptação às grandes altitudes em populações andinas. (Valverde *et al*., 2015).

Crawford e colaboradores (2017), mais recentemente, descreveram sinal de seleção em uma via até então não identificada em andinos. Os genes com evidências mais fortes de seleção - *BRINP3* (*Bone Morphogenetic Protein/Retinoic Acid Inducible Neural-Specific 3*), *NOS2* (*Nitric Oxide Synthase 2*) e *TBX5* (*T-Box 5*) - estão associados ao desenvolvimento e função cardiovascular, mas não atuam na via de resposta à hipóxia.

O aumento da quantidade de hemácias (policitemia) devido à hipóxia pode aumentar o risco de mortalidade por insuficiência cardíaca, causa comumente observada na doença crônica das montanhas (Mal das Montanhas). Além disso, as respostas fisiológicas à hipóxia estão associadas à redução de *fitness* adaptativo. Interessantemente, os tibetanos se adaptaram às grandes altitudes, em parte por meio da seleção de variantes genéticas associadas à redução da resposta fisiológica à policitemia. Em contraste, os nativos do Altiplano Andino apresentam uma elevada contagem de hemácias. Assim, é esperado que eles exibam outras características adaptativas que possam suprir os altos níveis de hemoglobina e células vermelhas. Por exemplo, haplótipos descritos para os três genes - citados previamente: *BRINP3*, *NOS2* e *TBX5* - podem estar associados às variações fenotípicas relacionadas à saúde cardiovascular dessas populações, a fim de mitigar os efeitos deletérios da policitemia em vez de reduzir a policitemia *per se* (Crawford *et al*., 2017)

Apesar de inúmeros genes e diferentes vias estarem implicadas na adaptação genética às altas altitudes, um número crescente de evidências sugere que apenas a genética não seria capaz de explicar a variação nos fenótipos adaptativos observados nessas populações. Para caracterizar se as respostas adaptativas nos andinos possuem um componente epigenético, Childebayeva e colaboradores avaliaram o padrão de metilação do DNA na região promotora do *EPAS1* e no elemento repetitivo *LINE-1* (*Long Interspersed Element-1*) em indivíduos Quechua do Altiplano Andino (4.388m acima do nível do mar) e indivíduos residentes de baixas altitudes do Peru. Nesse estudo, os autores

sugerem que os indivíduos nativos do Altiplano Andino (Quechua) têm um padrão de metilação do DNA menor no gene *EPAS1* e maior no *LINE-1*. Além disso, foram descritos quatro SNPs na via do metabolismo 1C (One Carbon) nos genes: *MTHFD1* (*Methylene Tetra Hydro Folate Dehydrogenase 1*) rs2236225, (*Thymidylate Synthetase*) rs502396,(*Folate Hydrolase 1*) rs202676,(*Glycine Decarboxylase*) rs1097568, os quais foram associados a diferenças na metilação de *LINE-1*. Os genes na via 1C estão envolvidos na geração de grupos metil para metilação do DNA, e, juntos explicam quase 12% da variação no padrão de metilação dos indivíduos que vivem em terras altas. Tomados em conjunto esses resultados indicam que a exposição à hipóxia afeta o padrão de metilação de *EPAS1* e *LINE*-1 nos indivíduos que vivem nas terras altas da América do Sul, sugerindo que modificações epigenéticas podem ter um papel importante na adaptação em altas altitudes (Childebayeva *et al*., 2019).

1.4 O gene *TP53* e sua *Network*

Um gene que recentemente vem se destacando na adaptação para vida em altas altitudes é o gene *TP53* (*Tumor Protein P53*), codificador da proteína p53. Eichstaedt e colaboradores (2015) foram os primeiros a reportarem a associação de *TP53* com a adaptação à altitude em populações nativas do Altiplano Andino. Através de abordagem GWAS, eles avaliaram a população de Colla (Andes, Argentina) e identificaram um sinal de seleção natural em diversos genes, inclusive no *TP53*. No entanto, à época, os autores apenas citaram o fato de *TP53* estar envolvido na via de resposta à hipóxia (Eichstaedt *et al*., 2015), sem dar o devide destaque a esse achado.

Em outros mamíferos não humanos, alterações mutacionais em *TP53* foram descritas como tendo um papel importante na adaptação ambiental. Zhao *et al*., (2013) compararam os genomas de roedores de terras altas (*Myospalax baileyi* e *Microtus oeconomus*) com roedores de terras baixas (*Myospalax cansus*), e observaram variações no códon 104 de p53 potencialmente funcionais. Estudos *in vitro* demonstraram como essas variantes respondem aos estímulos de hipóxia, frio e a regulação do pH. A presença de Ácido Glutâmico nesta posição em p53de *Microtus oeconomus* é relacionada à tolerância a hipóxia, enquanto a presença de uma Asparagina na mesma posição em *Myospalax baileyi* poderia ser uma adaptação tanto a hipóxia quanto a hipercapnia (estresse ácido, alto CO_2) relacionado às altas altitudes (~3,000–4,500 m) que essa espécie habita (Zhao *et al*., 2013).

Os resultados indicam que as variações de p53 observadas nestes animais de altitude não são aleatórias, mas provavelmente são produtos da seleção natural, o que explicaria parcialmente a sobrevivência às tensões ambientais características do planalto Tibetano (Zhao *et al*., 2013).

Conhecida como "guardiã do genoma", a proteína p53 funciona como um fator de transcrição, e desempenha um papel crítico nas respostas às diversas fontes estressoras a fim de manter a estabilidade genômica (**Figura 7**) (Chao, 2015). A quantidade de p53 nas células é determinada principalmente pela taxa na qual ela é degradada, ao invés da taxa pela qual é produzida. A proteína é constitutivamente expressa em todos as células nucleadas e é mantida em níveis basais devido à sua rápida degradação via proteassomo (Nag *et al*., 2013; Jovanović *et al*., 2005). No entanto, em resposta a diferentes sinais celulares ocorre um acúmulo de p53, além de modificações pós-transcricionais na proteína, que vão gerar uma resposta adequada a tais estímulos. Esses sinais incluem elementos que causam danos ao DNA (estresse genotóxico, como a radiação UV), estresse oncogênicos, depleção de ribonucleotídeos ou hipóxia (Chen, 2012; Yamamoto e Iwakuma, 2018).

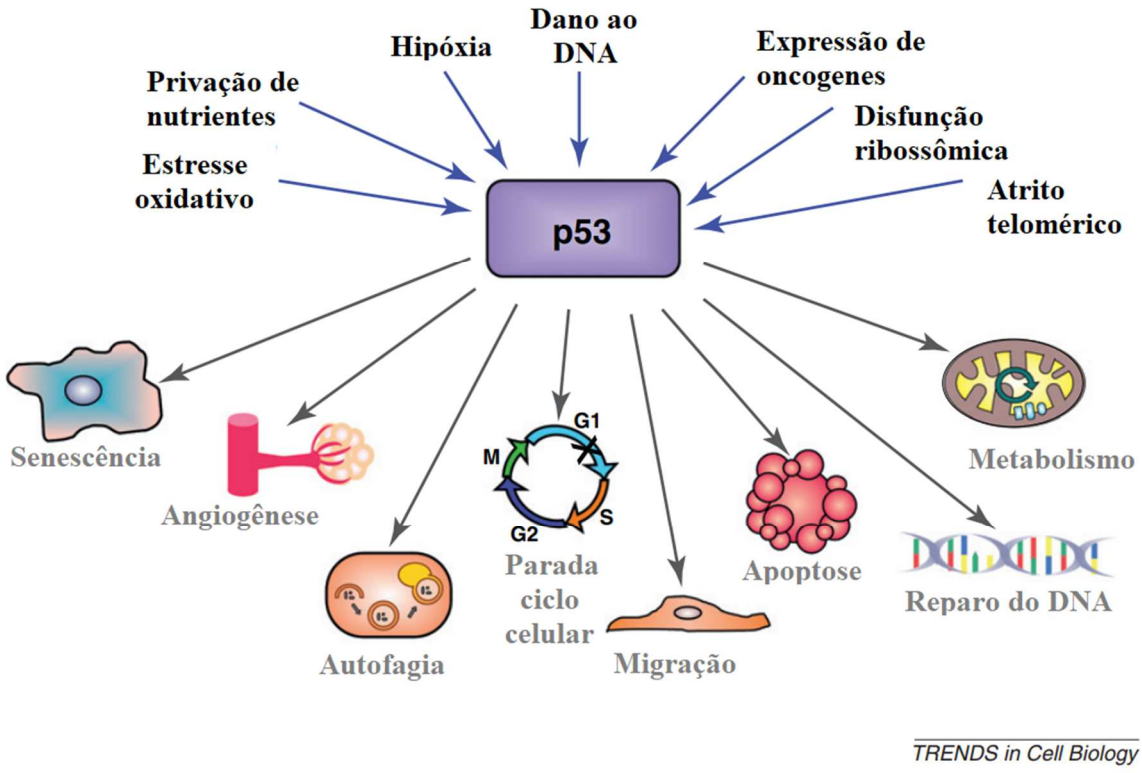


Figura 7. Mecanismos que podem ativar p53 e as possíveis respostas celulares desencadeadas. Adaptado e traduzido de Bieging e Attardi 2012.

A principal proteína que regula os níveis e atividade de p53 é a MDM2 (*Mouse Double Minute 2 Homolog*). MDM2 regula a atividade de p53 tanto no núcleo quanto no citoplasma. No núcleo ela é responsável por manter baixos níveis de p53 em homeostase normal. Em condições de estresse celular, MDM2 rapidamente ativa p53 através de mecanismos que bloqueiem a sua interação com p53, resultando num aumento da quantidade de p53 na célula. No citoplasma, MDM2 atua como uma E3-ubiquitina-ligase tornando p53 alvo da degradação pelo proteassomo. As proteínas p53 e MDM2 são reguladas através de um sistema de *feedback* autorregulado: p53 regula positivamente MDM2, estimulando a transcrição do gene, enquanto MDM2 regula negativamente p53 (Kohn e Pommier, 2005; Nag *et al*., 2013).

A resposta transcricional apropriada de p53 depende ainda de outros genes que compõem a via clássica de *TP53* (**Figura 8**), tais como *MDM2*, *MDM4*, *USP7* e *LIF*.

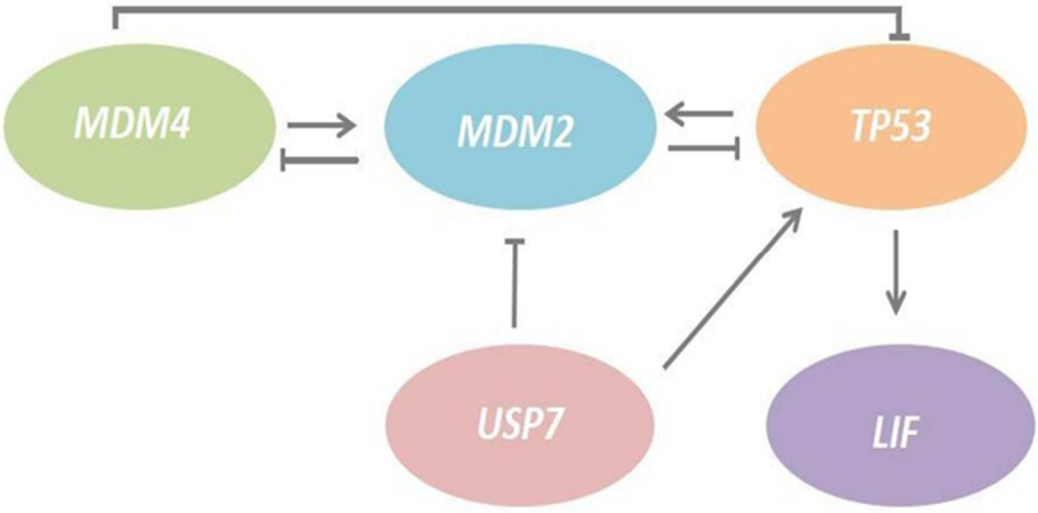


Figura 8. Rede clássica dos genes que interagem com *TP53* (Adaptado de Kang *et al*., 2009).

Entre as variantes polimórficas mais bem caracterizadas do gene *TP53*, uma em especial tem destaque: uma mutação não-sinônima que consiste na troca de uma prolina (Pro: alelo C) por uma arginina (Arg: alelo G) no códon 72 do éxon 4 (rs1042522). O Alelo Arg72 é mais ativo na indução de apoptose e supressão da transformação celular (Dumont *et al*., 2003).

O alelo Pro72 também é observado em outros primatas (incluindo o chimpanzé), enquanto o Arg72 está presente apenas em humanos, sugerindo que o alelo Pro72 seja o

alelo ancestral. Devido a funcionalidade dos diferentes alelos de p53 Pro72Arg e suas diferentes distribuições ao longo dos continentes, sugere-se que este fato não seria decorrente apenas de eventos aleatórios (Beckman *et al*., 1994 e Sjalander *et al*., 1996). Shi e colaboradores (2009) sugeriram que a ocorrência em altas frequências do alelo Arg72 em climas frios seria devido a ação da pressão seleção positiva, o que denota que dificilmente a variação hoje encontrada em *TP53,* bem como em outros genes da rede, não esteja sendo mediada pela ação seleção natural (Shi *et al*., 2009).

Recentemente, um estudo correlacionando a incidência de câncer ao redor do mundo e variáveis ambientais (tais como média de temperatura), sugeriu que a redução do potencial apoptótico seja ocasionado por variantes específicas na proteína p53, que são benéficas em ambientes de altitude, mas poderiam levar a suceptibilidade ao câncer nos dias atuais, pelo menos em populações andinas e tibetanas (Voskarides, 2018). Além disso, tem sido descrito o papel de genes da rede clássica de *TP53* na reprodução humana (Paskulin *et al*., 2012; Fraga *et al*., 2014), fator crítico para uma trajetória evolutiva de sucesso.

Estudos do polimorfismo no gene humano *MDM2* (T > G; rs2279744), por sua vez, indicam que o alelo derivado cria um sítio para o fator de transcrição Sp1. Como resultado, homozigotos para o alelo G expressam mais RNA mensageiro, e, consequentemente, níveis elevados de MDM2 são produzidos quando comparado aos homozigotos para o alelo T. Em situações de maior expressão de MDM2 há uma diminuição na expressão de p53 celular, o que diminuiria a resposta celular ao estresse, resultando em uma maior taxa de mutação, devido à prejuízos no processo de reparo do DNA. Tal redução na atividade apoptótica e senescência aumentaria o risco para tumorigênese (Atwal *et al*., 2007).

Vale destacar ainda que EPAS1 (HIF-2α) - que pertencente à família HIF - tem importante atuação na rede de *TP53* (Bertout *et al*., 2009). Os genes da via de HIF foram primeiramente descritos devido a sua conexão com células cancerosas, uma vez que HIF-1α induz a expressão do fator de crescimento endotelial vascular (VEGF) e eritropoietina, os quais promovem a angiogênese e a eritropoiese, respectivamente (Chen *et al*., 2003). A subunidade α de HIF (HIF-1α) regula a atividade de p53, através da estabilidade frente a danos celulares, além da sua exportação nuclear através de interações com MDM2 (Chen *et al*., 2003; Schmid *et al*., 2004; Zhou *et al*., 2015). Durante a hipóxia, há um acúmulo de HIF-1α. Em contrapartida, quando as células são reoxigenadas, HIF-1α é rapidamente

degradada via proteassomo, mediado por VHL. Deste modo, a proteína p53 só é aumentada sob hipóxia severa. Sendo assim, HIF desempenha um papel fundamental na proteção celular em hipóxia moderada (Fels e Koumenis, 2006).

Considerando o impacto de alterações da via de *TP53* na expressão de p53, bem como os estudos que indicam correlação entre essas alterações genéticas e variações climáticas (Kang *et al*., 2009; Shi *et al*., 2009; Sucheston *et al*., 2011; Zhao *et al*., 2013) somado ao papel de p53 na via de hipóxia, podemos pensar que variações gênicas nesta rota podem desempenhar um papel na adaptação humana mediada pelo clima, e assim ter um papel relevante nas adaptações a altas altitudes nos Andes.

Sendo assim, *TP53* e outros genes da mesma rede tornam-se estraordinários genes candidatos para estudos que buscam desvendar o repertório genético por trás da adaptação às altas altitudes. Este tema será adequadamente abordado e aprofundado no capítulo 3: "*Genetic Variations in the TP53 Pathway in Native Americans Strongly Suggest Adaptation to the High Altitudes of the Andes*" e capítulo 4: "*Selection Scan Reveals Three New Loci Related to High Altitude Adaptation in Native Andeans*".

1.5 O gene HLA-G

O Antígeno Leucocitário Humano G (HLA-G) é uma molécula não clássica do complexo principal de histocompatibilidade (MHC) de classe 1, que possui uma expressão restrita às células do trofoblasto extraviloso na interface materno-fetal durante a gestação. A expressão de HLA-G tem sido associada à tolerância imunológica materna em relação às células fetais, além de estimular a liberação de importantes fatores angiogênicos pelas células imunes maternas, sendo essencial para o sucesso da reprodução humana e manutenção da nossa espécie. É importante ressaltar que a expressão do HLA-G varia em condições patológicas e pode ter efeitos tanto prejudiciais (doenças infecciosas, neoplasias malignas) ou benéficos (autoimunidade e distúrbios inflamatórios) dependendo do contexto biológico (Alegre *et al*., 2014).

Notavelmente, a expressão de HLA-G é modulada *in vitro*, após a exposição à hipóxia. Além disso, os níveis de HLA-G são claramente relacionados à gestação e progressão de tumores, ambas condições caracterizadas por ambientes sob hipóxia (Nagamatsu *et al*., 2004; Yaghi *et al*., 2016).

A região 3’ UTR não traduzida do *HLA-G* desempenha um papel importante na regulação pós-transcricional do gene. Os polimorfismos de nucleotídeo único no *HLA-G* 3’UTR são relatados como influenciadores da expressão de HLA-G através de mecanismos relacionados à estabilidade do mRNA ou repressão traducional através da ligação de miRNA (Castelli *et al*., 2010).

Apesar do curto comprimento desta região [~358 pares de base (pb)], o *HLA-G* 3'UTR abriga inúmeras variantes. São elas:14-bp indel (rs371194629), +3001 C/T (rs567747015), +3003 T/C (rs1707), +3010 C/G (rs1710), +3027 C/A (rs17179101),+3032 G/C(rs146339774),+3035 C/T(rs17179108),+3052 C/T(rs569057854),+3092 G/T(rs180827037),+3107 C/G (rs não disponível),+3111 G/A(rs554784083),+3121 T/C(rs138249160),+3142 G/C(rs1063320),+3177 G/T(rs554076817),+3183A/G(rs187320344),+3187 A/G(rs9380142),+3196 C/G (rs1610696),+3227 G/A (rs1233331),+3289 C/T(rs541542414). Até o momento, diversos haplótipos já foram caracterizados através da análise destes sítios polimórficos em diferentes populações. No entanto, apenas 8 haplótipos são observados em uma frequência acima de 1% (UTR-1, -2, -3, -4, -5. -6, -7, -18). Além disso, a UTR-1, a UTR-2 e a UTR-3 são as mais frequentes e representam cerca de 70% da variação observada em todo o mundo (Sabbagh *et al*., 2014).

É descrito que os haplótipos de HLA-G 3'UTR possam influenciar os níveis de HLA-G solúvel (sHLA-G). O haplótipo UTR-1 geralmente é associado a níveis plasmáticos mais elevados de sHLA-G, uma vez que é constituído pela maioria das variantes independentemente associadas à maior expressão proteica. Por outro lado, a UTR-5 e a UTR-7 estariam associados aos níveis mais baixos de sHLA-G, enquanto UTR-2, 3, 4 e 6 exibiram níveis intermediários solúveis (Martelli-Palomino *et al*., 2013).

No contexto da alta altitude, avaliando o sangue periférico de alpinistas no Monte Everest (planalto do Tibete) Bourguignon e colaboradores (2010) evidenciaram que os níveis de sHLA-G são regulados positivamente durante a escalada até o cume da montanha, indicando que HLA-G pode ser um mecanismo relevante na adaptação as altas altitudes, pelo menos considerando adaptação fisiológica, transiente e reversível (*short-term*). Os autores também reforçaram o papel da hipóxia como um fator importante na regulação da expressão de HLA-G (Bourguignon *et al*., 2010).

A região *HLA-G* 3'UTR foi caracterizada na população brasileira do estado de São Paulo (Castelli *et al*., 2010). Neste estudo, 8 polimorfismos foram descritos - caracterizando 7 haplótipos, UTR-1 a UTR-7. Posteriormente, outros grupos populacionais brasileiros foram explorados, indicando que o padrão da diversidade dos haplótipos da região 3'UTR do *HLA-G* observado em São Paulo também é observado em outras regiões do país (Mendes *et al*., 2007; Lucena-Silva *et al*., 2012). A variação genética da região *HLA-G* 3'UTR já foi estudada na Europa, África Subsaariana, Ásia Oriental e América (populações miscigenadas), avaliadas em um total de 21 populações. Desse modo, a diversidade e a caracterização dessa região gênica em Nativos Americanos, e, na adaptação à alta altitude ainda é desconhecida.

Evolutivamente, Castelli e colaboradores (2011) descreveram evidências de seleção purificadora atuando sobre a região codificadora do *HLA-G* (Castelli *et al*., 2011), enquanto que as duas regiões regulatórias, 5'-UTR e 3'-UTR, apresentam evidências de seleção balanceadora em diferentes populações (Tan *et al*., 2005; Mendes *et al*., 2007; Castelli *et al*., 2011; Donadi *et al*., 2011; Veit *et al*., 2012; Mendes-Junior *et al*., 2013).

Tomados em conjunto, esses dados sugerem que o *HLA-G* pode ser um gene candidato bastante interessante para ser avaliado no contexto da adaptação para a vida em altitude no Altiplano Andino. Desta forma, no capítulo 5 é apresentado através do manuscrito, em preparação, intitulado "*Characterization of HLA-G 3'UTR in Native Americans– altitude as a selection factor influencing population frequencies of HLA-G haplotypes*".

1.6 A Busca por Genes e suas variantes através das varreduras genômicas

Como visto nos itens anteriores, parte dos estudos que identificaram até aqui as bases genéticas por trás da adaptação humana as altas altitudes se deu por uso das estratégias de genes candidatos. Nesse caso, concentra-se em associações entre a variação genética em gene(s) pré-especificado(s) e com já conhecido (ou sugerido) papel funcional nos fenótipos sob investigação (Zhu e Zhao, 2007). Esse tipo de abordagem se contrapõe aos estudos de associação com dados genéticos de varredura genômica (*genome-wide association studies*, GWAS), que usa como estratégia um grande conjunto de polimorfismos comuns, para identificar distribuições alélicas significantemente diferentes entre os conjuntos populacionais investigados. Dessa forma, novos genes e variantes

podem ser identificados como membros de um repertório genético associado a um determinado fenótipo. Tal estratégia foi basicamente desenvolvida para se identificar fatores de risco genéticos para doenças comuns e complexas, como esquizofrenia e diabetes tipo II (Bush *et al*., 2012). Não obstante, estudos que usam GWAS vêm sendo útil também na identificação do conjunto de variantes comuns que podem estar relacionados com traços adaptativos.

Já foi visto em itens anteriores várias referências à estudos de GWAS que são pertinentes a temática da presente tese. Usamos essa estratégia para desenvolver um trabalho que faz parte da presente tese e que já foi publicado:"*Selection scan reveals three new loci related to high altitude adaptation in Native Andeans*".

1.7 Referências Bibliográficas

Acuña-Alonzo V, Flores-Dorantes T, Kruit JK, Villarreal-Molina T, Arellano-Campos O, Hünemeier T, Moreno-Estrada A, Ortiz-López MG, Villamil-Ramírez H, León-Mimila P *et al*., (2010) A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. Hum Mol Genet 19:2877-2885.

Aldenderfer M (2011) Peopling the Tibetan plateau: insights from archaeology. High Alt Med Biol 12:141-147.

Alegre E, Rizzo R, Bortolotti D, Fernandez-Landazuri S, Fainardi E, Gonzalez A, (2014) Some basic aspects of HLA-G biology. J Immunol Res 2014:657625.

Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A (2012) The genetic architecture of adaptations to high altitude in Ethiopia. PLoS Genet 8:e1003110.

Amorim CE, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, Hünemeier T (2017) Genetic signature of natural selection in first Americans. Proc Natl Acad Sci USA 114:2195-2199.

Atwal GS, Kirchhoff T, Bond EE, Montagna M, Menin C, Bertorelle R, Scaini MC, Bartel F, Böhnke A, Pempe C *et al*., (2007) Haplotype structure and selection of the MDM2 oncogene in humans. Proc Natl Acad Sci U S A 104:4524-4529.

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M *et al*., (2010) Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci U S A 107:11459-11464.

Beall CM, Decker MJ, Brittenham GM, Kushner I, Gebremedhin A, Strohl KP (2002) An Ethiopian pattern of human adaptation to high-altitude hypoxia. Proc Natl Acad Sci USA 99:17215-17218.

Beall CM, Laskowski D, Erzurum SC (2012) Nitric oxide in adaptation to altitude. Free Radic Biol Med 52:1123-1134.

Beckman G, Birgander R, Själander A, Saha N, Holmberg PA, Kivelä A, Beckman L (1994) Is p53 polymorphism maintained by natural selection? Hum Hered 44:266–270.

Bieging KT, Attardi LD (2012) Deconstructing p53 transcriptional networks in tumor suppression. Trends Cell Biol 22(2):97-106.

Bertout JA, Majmundar AJ, Gordan JD, Lam JC, Ditsworth D, Keith B, Brown EJ, Nathanson KL, Simon MC (2009) HIF2alpha inhibition promotes p53 pathway activity, tumor cell death, and radiation responses. Proc Natl Acad Sci U S A 106(34):14391-6.

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, López Herráez D (2010) Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet 6:e1001116.

Bigham AW, Julian CG, Wilson MJ, Vargas E, Browne VA, Shriver MD, Moore LG (2014) Maternal PRKAA1 and EDNRA genotypes are associated with birth weight, and PRKAA1 with uterine artery diameter and metabolic homeostasis at high altitude. Physiol Genomics 46:687-697.

Bigham AW and Lee FS (2014) Human high-altitude adaptation: forward genetics meets the HIF pathway. Genes Dev 28:2189-2204.

Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, Julian CG, Parra EJ, Akey JM, Moore LG, Shriver MD (2009) Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Hum Genomics 4:79-90.

Bigham AW (2016) Genetics of human origin and evolution: high-altitude adaptations. Curr Opin Genet Dev 41:8-13.

Borradale DC and Kimlin MG (2012) Folate degradation due to ultraviolet radiation: Possible implications for human health and nutrition. Nutr Rev 70:414-422.

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 111: E213-214.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT *et al*., (2003) Y-Chromosome for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524-539.

Bourguignon M, Yaghi L, Flajollet S, Radanne-Krawice I, Rouas-Freiss N, Lugrin D, Richalet JP, Carosella ED, Moreau P, (2010) Increased soluble human leukocyte antigen-G levels in peripheral blood from climbers on Mount Everest. Hum Immunol 71: 1105–1108.

Bush WS, Moore JH (2012) Lewitter F, Kann M, eds. "Chapter 11: Genome-wide association studies". PLoS Computational Biology 8 (12): e1002822.

Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, Roger M, Moreau P, Donadi EA (2011) A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. Mol Biol Evol 28:3069-86.

Castelli EC, Mendes-Junior CT, Deghaide NH, de Albuquerque RS, Muniz YC, Simões RT, Carosella ED, Moreau P, Donadi EA (2010) The genetic structure of 3'untranslated region of the HLA-G gene: polymorphisms and haplotypes. Genes Immun 11:134-141.

Chala-Aldana D, Bocherens H, Miller C, Moore K, Hodgins G, Rademaker K (2017) Investigating mobility and highland occupation strategies during the Early Holocene at the Cuncaicha rock shelter through strontium and oxygen isotopes. J Archaeol Sci Rep 19:811-827.

Chao CC (2015) Mechanisms of p53 degradation. Clin Chim Acta 438:139-147.

Chen D, Li M, Luo J, Gu W (2003) Direct interactions between HIF-1 alpha and Mdm2 modulate p53 function. J Biol Chem 278:13595–13598.

Chen J (2012) The Roles of MDM2 and MDMX Phosphorylation in Stress Signaling to p53. Genes Cancer 3:274-282.

Childebayeva A, Jones TR, Goodrich JM, Leon-Velarde F, Rivera-Chira M, Kiyamu M, Brutsaert TD, Dolinoy DC, Bigham AW (2019) LINE-1 and EPAS1 DNA Methylation Associations with High-Altitude Exposure. Epigenetics 14(1):1-15.

Crawford JE, Ricardo Amaru R, Song J, Julian CG, Racimo F, Cheng JY, Guo X, Yao J, Ambale-Venkatesh B, Lima JA (2017) Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. Am J Hum Genet 101:752-767.

Cueto M (1986) Andean biology in Peru: scientific styles on the periphery. Isis 80: 640–658.

Donadi EA, Castelli EC, Arnaiz-Villena A, Roger M, Rey D, Moreau P (2011) Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. Cell Mol Life Sci 68:369-395.

Dumont P, Leu JI, Della Pietra AC 3rd, George DL, Murphy M (2003) The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet 33:357–365.

Eichstaedt CA, Antão T, Pagani L, Cardona A, Kivisild T, Mormina M (2014) The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS One 9:e93314.

Eichstaedt CA, Pagani L, Antao T, Inchley CE, Cardona A, Mörseburg A, Clemente FJ, Sluckin TJ, Metspalu E, Mitt M *et al*., (2017) Evidence of Early-Stage Selection on EPAS1 and GPR126 Genes in Andean High Altitude Populations. Sci Rep 7:13042.

Eichstaedt CA, Antao T, Cardona A, Pagani L, Kivisild T, Mormina M (2015) Positive selection of AS3MT to arsenic water in Andean populations. Mutat Res Mol Mech Mutagen 780:97-102.

Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Silva Jr WA, Zago MA, Ribeiro-dos-Santos AK *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 82:583-592.

Fan S, Hansen ME, Lo Y, Tishkoff SA (2016) Going global by adapting local: A review of recent human adaptation. Science 354: 54-59.

Fehren-Schmitz L and Georges L (2016) Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian Highlanders in the last 8500 years. Sci Rep 6:23485.

Fels DR and Koumenis C (2006) The PERK/eIF2alpha/ATF4 module of the UPR in hypoxia resistance and tumor growth. Cancer Biol Ther 5(7):723-8.

Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L (2014) Widespread signals of convergent adaptation to high altitude in Asia and America. Am J Hum Genet 95:394-407.

Fraga LR, Dutra CG, Boquett JA, Vianna FS, Gonçalves RO, Paskulin DD, *et al*., (2014) p53 signaling pathway polymorphisms associated to recurrent pregnancy loss. Mol Biol Rep 41:1871-1877.

Frisancho AR (2013) Developmental functional adaptation to high altitude: Review. Am J Hum Biol 25:151-168.

Fort M, de Stefano GF, Cambon-Thomsen A, Giraldo-Alvarez P, Dugoujon JM, Ohayon E, Scano G, Abbal M (1998) HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. Tissue Antigens 1:327-336.

Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, Korneliussen TS, Gerbault P, Skotte L, Linneberg A *et al*., (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science 349:1343-1347.

González-José R, Bortolini MC, Santos FR and Bonatto SL (2008) The peopling of America: craniofacial shape variation on a continental scale and its interpretation from na interdisciplinary view. Am J Phys Anthropol 137:175-187.

Greer SN, Metcalf JL, Wang Y, Ohh M (2012) The updated biology of hypoxia-inducible factor. EMBO J 31:2448-2460.

Henry JP (2019) Genetics and origin of Homo sapiens. Med Sci Paris 35:39-45.

Hu H, Petousi N, Glusman G, Yu Y, Bohlender R, Tashi T, Downie JM, Roach JC, Cole AM, Lorenzo FR *et al*., (2017) Evolutionary history of Tibetans inferred from whole-genome sequencing. PLoS Genet 13:e1006675.

Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, Bergmann I, Le Cabec A, Benazzi S, Harvati K *et al*., (2017) New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature 546:289-292.

Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Nielsen R *et al.,* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 14;512:194-7.

Hünemeier T, Amorim CE, Azevedo S, Contini V, Acuña-Alonzo V, Rothhammer F, Dugoujon JM, Mazières S, Barrantes R, Villarreal-Molina MT *et al*., (2012) Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. PLoS One 7(6):e38862.

Jeong C, Witonsky DB, Basnyat B, Neupane M, Beall CM, Childs G, Craig SR, Novembre J, Di Rienzo A (2018) Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. PLoS Genet14:e1007650.

Jovanović I, Todorović V, Milosavljević T, Micev M, Pesko P, Bjelović M, Mouzas Y, Tzardi M (2005) Expression of p53 protein in Barrett's adenocarcinoma and adenocarcinoma of the gastric cardia and antrum. Vojnosanit Pregl 62:879-885.

Kang HJ, Feng Z, Sun Y, Atwal G, Murphy ME, Rebbeck TR, Rosenwaks Z, Levine AJ, Hu W (2009) Single-nucleotide polymorphisms in the p53 pathway regulate fertility in humans. Proc Natl Acad Sci U S A106:9761-9766.

Keith B, Johnson RS, Simon MC (2011) HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression. Nat Rev Cancer12:9-22.

Kohn KW and Pommier Y (2005) Molecular interaction map of the p53 and Mdm2 logic elements, which control the Off-On switch of p53 in response to DNA damage. Biochem Biophys Res Commun 331816-827.

Lanikova L, Reading NS, Hu H, Tashi T, Burjanivova T, Shestakova A, Siwakoti B, Thakur BK, Pun CB, Sapkota A *et al*., (2017) Evolutionary selected Tibetan variants of HIF pathway and risk of lung cancer. Oncotarget 8:11739-11747.

Lee FS and Percy MJ (2011) The HIF pathway and erythrocytosis. Annu Rev Pathol 6:165-92.

Lindo J, Haas R, Hofman C, Apata M, Moraga M, VerdugozRA, Watson JT, Viviano Llave C, Witonsky D, Beall C *et al*., (2018) The genetic prehistory of the Andean highlands 7000 years BP though European contact. Sci Adv 4:eaau4921.

Lorenzo FR, Huff C, Myllymäki M, Olenchock B, Swierczek S, Tashi T, Gordeuk V, Wuren T, Ri-Li G, McClain DA, *et al.,* (2014) A genetic mechanism for Tibetan high-altitude adaptation. Nat Genet. 46:951-6.

Lou H, Lu Y, Lu D, Fu R, Wang X, Feng Q, Wu S, Yang Y, Li S, Kang L *et al*., (2015) A 3.4-kb copy-number deletion near EPAS1 is significantly enriched in high-altitude Tibetans but absent from the Denisovan sequence. Am J Hum Genet 97:54-66.

Lucena-Silva N, Monteiro AR, de Albuquerque RS, Gomes RG, Mendes-Junior CT, Castelli EC, Donadi EA (2012) Haplotype frequencies based on eight polymorphic sites at the 3' untranslated region of the HLA-G gene in individuals from two different geographical regions of Brazil. Tissue Antigens 79(4):272-278.

Martelli-Palomino, G., Pancotto, J.A., Muniz, Y.C., Mendes-Junior, C.T., Castelli, E.C., Massaro, J.D., Krawice-Radanne, I., Poras, I., Rebmann, V., Carosella, E.D *et al.,* (2013) Polymorphic Sites at the 3' Untranslated Region of the HLA-G Gene Are Associated with Differential HLA-G Soluble Levels in the Brazilian and French Population. PLoS One 8, e71742.

Monge CM (1948) Acclimatization in the Andes: Historical Confirmations of 'Climatic Aggression' in the Development of Andean Man. Johns Hopkins Press, 1948.

Mendes-Junior CT, Castelli EC, Simões RT, Simões AL, Donadi EA (2007) HLA-G 14-bp polymorphism at exon 8 in Amerindian populations from the Brazilian Amazon. Tissue Antigens 69:255-260.

Miyazaki K, Kawamoto T, Tanimoto K, Nishiyama M, Honda H, Kato Y (2002) Identification of functional hypoxia response elements in the promoter region of the DEC1 and DEC2 genes. J Biol Chem 277:47014-47021.

Moore LG (2001) Human genetic adaptation to high altitude. High Alt Med Biol. 2:257-79.

Moore LG (2017) Measuring high-altitude adaptation. J Appl Physiol 123:1371-1385.

Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, Kamm JA, Albrechtsen A, Malaspinas AS, Sikora M *et al*., (2018) Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature 553:203-207.

Nag S, Qin J, Srivenugopal KS, Wang M, Zhang R (2013) The MDM2-p53 pathway revisited. J Biomed Res 27:254-271.

Nagamatsu T, Fujii T, Yamashita T, Miki A, Kanai T, Kusumi M, Osuga Y, Kozuma S, Taketani Y (2004) Hypoxia does not reduce HLA-G expression on extravillous cytotrophoblasts. J Reprod Immunol 63:85-95.

Paskulin, Paixão-Côrtes VR, Hainaut P, Bortolini MC, Ashton-Prolla P (2012) The TP53 fertility network. Genet Mol Biol 35:939-946.

Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B and Oyunsuren T (1995) A major founder Y-chromosome haplotype in Amerindians. Nat Genet 11:15-16.

Peng Y, Cui C, He Y, Ouzhuluobu, Zhang H, Yang D, Zhang Q, Bianbazhuoma, Yang L, He Y *et al.,* (2017) Down-Regulation of EPAS1 Transcription and Genetic Adaptation of Tibetans to High-Altitude Hypoxia. Mol Biol Evol 34:818-830.

Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu Basang *et al*., (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol 28:1075-1081.

Pinotti T, Bergström A, Geppert M, Bawn M, Ohasi D, Shi W, Lacerda DR, Solli A, Norstedt J, Reed K *et al*., (2019) Y Chromosome Sequences Reveal a Short Beringian Standstill, Rapid Expansion, and early Population structure of Native American Founders. Curr Biol 29:149-157.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 175:1185-1197.e22.

Rademaker K, Hodgins G, Moore K, Zarrillo S, Miller C, Bromley GR, Leach P, Reid DA, Álvarez WY, Sandweiss DH (2014) Paleoindian settlement of the high-altitude Peruvian Andes. Science. 24;346:466-9.

Ray N, Wegmann D, Fagundes NJ, Wang S, Ruiz-Linares A and Excoffier L (2009) A statistical evaluation of models for the initial settlement of the American continente emphasizes the importance of gene flow with Asia. Mol Biol Evol 27:337-345.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N *et al*., (2012) Reconstructing Native American population history. Nature 488:370-374.

Sabbagh A, Luisi P, Castelli EC, Gineau L, Courtin D, Milet J, Massaro JD, Laayouni H, Moreau P, Donadi EA *et al*., (2014) Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity. Genes Immun 15:95-106.

Santos FR, Bonatto SL and Bortolini MC (2007) Molecular evidence from contemporary indigenous populations to the Peopling of America. Recent advances in molecular biology and evolution: Applications to biological anthropology. Research Signpost, Kerala, pp 1-13.

Scheinfeldt LB, Soi S, Thompson S, Ranciaro A, Woldemeskel D, Beggs W, Lambert C, Jarvis JP, Abate D, Belay G *et al*., (2012) Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol 13: R1.

Scheinfeldt LB, Soi S and Tishkoff SA (2010) Colloquium paper: working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. Proc Natl Acad Sci U S A 2:8931-8938.

Scheinfeldt LB and Tishkoff SA (2010) Living the high life: high-altitude adaptation. Genome Biol 11:133.

Schlebusch CM, Gattepaille LM, Engström K, Vahter M, Jakobsson M, Broberg K (2015) Human adaptation to arsenic-rich environments. Mol Biol Evol 32:1544-55.

Schmid T, Zhou J and Brüne, B (2004) HIF-1 and p53: communication of transcription factors under hypoxia. J Cell Mol Med 8:423-431.

Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI and Smith DG (2007) A private allele ubiquitous in the Americas. Biol Lett 3:218-223.

Scortegagna M, Ding K, Zhang Q, Oktay Y, Bennett MJ, Bennett M, Shelton JM, Richardson JA, Moe O, Garcia JA (2005) HIF-2alpha regulates murine hematopoietic development in an erythropoietin-dependent manner. Blood 105:3133-3140.

Semenza GL (1999) Regulation of mammalian O2 homeostasis by hypoxia-inducible Factor 1. Annu Rev Cell Dev Biol 15:551-578.

Shi H, Tan SJ, Zhong H, Hu W, Levine A, Xiao CJ, Peng Y, Qi XB, Shou WH, Ma RL *et al.,* (2009) Winter temperature and UV are tightly linked to genetic changes in the p53 tumor suppressor pathway in Eastern Asia. Am J Hum Genet 84:534-541.

Shimoda LA and Laurie SS (2014) HIF and pulmonary vascular responses to hypoxia. J Appl Physiol 1:116:867-74.

Simonson TS, Huff CD, Witherspoon DJ, Prchal JT, Jorde LB (2015) Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans. Exp Physiol 100:1263-1268.

Simonson TS, McClain DA, Jorde LB, Prchal JT (2012) Genetic determinants of Tibetan high-altitude adaptation. Hum Genet 131:527-533.

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB *et al.,* (2010) Genetic evidence for high-altitude adaptation in Tibet. Science. 329:72-5.

Simonson TS, Huff C D, Witherspoon DJ, Prchal JT, Jorde LB (2015) Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans. Exp Physiol 100:1263-1268.

Själander A, Birgander R, Saha N, Beckman L, Beckman G (1996) p53 polymorphisms and haplotypes show distinct differences between major ethnic groups. Hum Hered 46:41–48.

Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, Salzano FM, Patterson N, Reich D (2015) Genetic evidence for two founding populations of the Americas. Nature 525:104-108.

Song D, Li LS, Arsenault PR, Tan Q, Bigham AW, Heaton-Johnson KJ, Master SR, Lee FS (2014) Defective Tibetan PHD2 binding to p23 links high altitude adaption to altered oxygen sensing. The Journal of biological chemistry 289:14656-14665.

Soria R, Julian CG, Vargas E, Moore LG, Giussani DA (2013) Graduated effects of high-altitude hypoxia and highland ancestry on birth size. Pediatr Res 74:633-638.

Sucheston L, Witonsky DB, Hastings D, Yildiz O, Clark VJ, Di Rienzo A, Onel K (2011) Natural selection and functional genetic variation in the p53 pathway. Hum Mol Genet 20:1502-1508.

Tan Z, Shon A, Ober C (2005) Evidence of balancing selection at the HLA-G promoter region. Hum Mol Genet 14:3619-3628.

Tashi T, Scott Reading N, Wuren T, Zhang X, Moore LG, Hu H, Tang F, Shestakova A, Lorenzo F, Burjanivova T *et al.,* (2017) Gain-of-function EGLN1 prolyl hydroxylase (PHD2 D4E:C127S) in combination with EPAS1 (HIF-2α) polymorphism lowers hemoglobin concentration in Tibetan highlanders. J Mol Med 95:665-670.

Valverde G, Zhou H, Lippold S, de Filippo C, Tang K, López Herráez D, Li J, Stoneking M (2015) A novel candidate region for genetic adaptation to high altitude in Andean populations. PLoS One 10:e0125444.

Veit TD, Cazarolli J, Salzano FM, Schiengold M, Chies JA (2012) New evidence for balancing selection at the HLA-G locus in South Amerindians. Genet Mol Biol 35:919-923.

Voskarides K (2018) Combination of 247 Genome-Wide Association Studies Reveals High Cancer Risk as a Result of Evolutionary Adaptation. Mol Biol Evol 35:473-485.

Wang B, Zhang YB, Zhang F, Lin H, Wang X, Wan N, Ye Z, Weng H, Zhang L, Li X *et al*., (2011) On the origin of Tibetans and their genetic basis in adapting high-altitude environments. PLoS One 6:e17002.

Wang GL and Semenza GL (1995) Purification and characterization of hypoxia-inducible factor 1. J Biol Chem 270:1230-1237.

Wang S, Lewis Jr CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C *et al*., (2007) Genetic variation and population structure in Native Americans. PLoS Genet 3:e185.

Winslow RM and Monge C (1987) Hypoxia, Polycythemia, and Chronic Mountain Sickness, Johns Hopkins Univ Press, Baltimore.

Woyessa A, Deressa W, Ali A, Lindtjørn B (2013) Malaria risk factors in Butajira area, south-central Ethiopia: a multilevel analysis. Malar J 12:273

Yaghi L, Poras I, Simoes RT, Donadi EA, Tost J, Daunay A, de Almeida BS, Carosella ED, Moreau P (2016) Hypoxia inducible factor-1 mediates the expression of the immune checkpoint HLA-G in glioma cells through hypoxia response element located in exon 2. Oncotarget 7: 1-18.

Yamamoto S and Iwakuma T (2018) Regulators of Oncogenic Mutant TP53 Gain of Function. Cancers (Basel) 20:11.

Yang D, Peng Y, Ouzhuluobu, Bianbazhuoma, Cui C, Bianba, Wang L, Xiang K, He Y, Zhang H *et al.,* (2016) HMOX2 Functions as a Modifier Gene for High-Altitude Adaptation in Tibetans. Hum Mutat 37:216-223.

Yang J, Jin ZB, Chen J, Huang XF, Li XM, Liang YB, Mao JY, Chen X, Zheng Z, Bakshi A *et al*., (2017) Genetic signatures of high-altitude adaptation in Tibetans. Proc Natl Acad Sci U S A 114:4189-4194.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS*, et al.,* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329, 75–78.

Zhao Y, Ren JL, Wang MY, Zhang ST, Liu Y, Li M, Cao YB, Zu HY, Chen XC, Wu CI., *et al*., 2013. Codon 104 variation of p53 gene provides adaptive apoptotic responses to extreme environments in mammals of the Tibet plateau. Proc Natl Acad Sci U S A 11051:20639-20644.

Zhou CH, Zhang XP, Liu F, Wang W (2015) Modeling the interplay between the HIF-1 and p53 pathways in hypoxia. Sci Rep 5:13834.

Zhu M and Zhao S (2007) Candidate gene identification approach: progress and challenges. Int J Biol Sci 3 (7): 420–427.

Internet resources section

Atlas of Genetics and Cytogenetics in Oncology and Haematology, http://atlasgeneticsoncology.org/ (January, 2019).

Altitude: http://www.altitude.org/home.php (January,2019).

Center for International Earth Science Information Network (CIESIN), http://www.ciesin.org/ (January, 2019).

CAPÍTULO 2

OBJETIVOS, MÉTODOS E RESULTADOS

2.1 Objetivo Geral

Avaliar o papel de variantes genéticas em populações ameríndias sul americanas, oriundas de ecossistemas distintos, considerando inicialmente a comparação de nativos americanos que habitam duas regiões geográficas distintas, quanto a vários parâmetros climáticos: terras altas, Andes (acima de 2.500 metros do nível do mar) e terras baixas (abaixo de 500 metros). Com isso, buscou-se inferir se as variantes (ou uma combinação delas), bem como o padrão geral de diversidade encontrados, considerando os *loci* sob investigação, poderiam ser explicados pela ação da seleção natural.

2.2 Objetivos Específicos

- Estimar a variabilidade de cinco genes da rede clássica de *TP53* através da investigação dos SNPs: rs1042522 (*TP53*), rs929271 (*LIF*), rs2279744 (*MDM2*), rs1563828 (*MDM4*), rs15229916 (*USP7*) nas populações alvo do estudo;
- Estimar a diversidade da região promotora 3'UTR do gene *HLA-G* nas populações alvo do estudo;
- A partir de uma varredura genômica, identificar genes e variantes genéticas associadas a adaptação a altas altitudes, considerando as populações alvo do estudo;
- Correlacionar as frequências genotípicas e alélicas encontradasnas três abordagensda presente tese (genes candidados da rede do *TP53*, região candidata 3'UTR do *HLA-G*e GWAS)nasdiferentes populações alvo com as variáveis geográficas, ambientais e climáticas;
- Identificar padrões e sinais que possam ser explicados pela ação da seleção natural;
- Montar um cenário evolutivo que explique os achados considerando a história evolutiva e demográfica das populações investigadas.

2.3. Métodos

Os métodos específicos para atingir os objetivos descritos acima, encontram-se descritos, em detalhes, na sessão de resultados considerando cada um dos artigos e o manuscrito(Capítulos **3**, **4** e **5**).

2.4. RESULTADOS

Nos capítulos 3, 4 e 5 pode-se encontrar os resultados da presente ese. Os mesmos foram reunidos em dois artigos (já publicados) e um manuscrito em preparação.

Além disso, um capítulo especial (6), relativo ao desenvolvido de um trabalho específico durante o transcurso desse doutorado, também é apresentado.

CAPÍTULO 3

Genetic Variations in the TP53 Pathway in Native Americans Strongly Suggest Adaptation to the High Altitudes of the Andes.

Publicado como um artigo científico: **Jacovas VC,** Rovaris DL, Peréz O, de Azevedo S, Macedo GS, Sandoval JR, Salazar-Granara A, Villena M, Dugoujon JM, Bisso-Machado R, Petzl-Erler ML, Salzano FM, Ashton-Prolla P, Ramallo V, Bortolini MC. *Genetic Variations in the TP53 Pathway in Native Americans Strongly Suggest Adaptation to the High Altitudes of the Andes.* PLoS One. 2015 Sep 8;10(9):e0137823. doi: 10.1371/journal.pone.0137823. eCollection 2015. PubMed PMID: 26382048; PubMed Central PMCID: PMC4575214.

PLOS ONE

RESEARCH ARTICLE

Genetic Variations in the *TP53* Pathway in Native Americans Strongly Suggest Adaptation to the High Altitudes of the Andes

Vanessa Cristina Jacovas[1☯], Diego Luiz Rovaris[1☯], Orlando Peréz[2], Soledad de Azevedo[2], Gabriel Souza Macedo[1], José Raul Sandoval[3], Alberto Salazar-Granara[3], Mercedes Villena[4], Jean-Michel Dugoujon[5], Rafael Bisso-Machado[1], Maria Luiza Petzl-Erler[6], Francisco Mauro Salzano[1], Patricia Ashton-Prolla[1,7], Virginia Ramallo[2], Maria Cátira Bortolini[1]*

1 Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil, **2** Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Puerto Madryn, Argentina, **3** Facultad de Medicina Humana, Universidad de San Martin de Porres (USMP), Lima, Peru, **4** Instituto Boliviano de Biología de Altura (IBBA), Universidad Mayor de San Andres, La Paz, Bolivia, **5** Anthropologie Moléculaire et Imagerie de Synthèse, CNRS UMR 5288, Université Paul Sabatier Toulouse III, Toulouse, 31000, France, **6** Laboratório de Genética Molecular, Departamento de Genética, Universidade Federal do Paraná, Curitiba, PR, Brasil, **7** Serviço de Genética Medica, Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil

☯ These authors contributed equally to this work.
* maria.bortolini@ufrgs.br

Citation: Jacovas VC, Rovaris DL, Peréz O, de Azevedo S, Macedo GS, Sandoval JR, 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(9): e0137823. doi:10.1371/journal.pone.0137823

Editor: Klaus Roemer, University of Saarland Medical School, GERMANY

Received: July 3, 2015

Accepted: August 24, 2015

Published: September 18, 2015

Data Availability Statement: All data are available in the manuscript and supporting information files.

Funding: This work was supported the by Conselho Nacional de Desenvolvimento Científico e Tecnológico, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and Fundo de Incentivo à Pesquisa e Eventos do Hospital de Clínicas de Porto Alegre (Brazil).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

The diversity of the five single nucleotide polymorphisms located in genes of the *TP53* pathway (*TP53*, rs1042522; *MDM2*, rs2279744; *MDM4*, rs1563828; *USP7*, rs1529916; and *LIF*, rs929271) were studied in a total of 282 individuals belonging to Quechua, Aymara, Chivay, Cabanaconde, Yanke, Taquile, Amantani, Anapia, Uros, Guarani Ñandeva, and Guarani Kaiowá populations, characterized as Native American or as having a high level (> 90%) of Native American ancestry. In addition, published data pertaining to 100 persons from five other Native American populations (Surui, Karitiana, Maya, Pima, and Piapoco) were analyzed. The populations were classified as living in high altitude ($\geq$ 2,500 m) or in lowlands (< 2,500 m). Our analyses revealed that alleles *USP7*-G, *LIF*-T, and *MDM2*-T showed significant evidence that they were selected for in relation to harsh environmental variables related to high altitudes. Our results show for the first time that alleles of classical *TP53* network genes have been evolutionary co-opted for the successful human colonization of the Andes.

Introduction

The product of the *TP53* gene is a transcription factor (p53) that activates or represses a large number of target genes that regulate a broad array of extremely important cellular functions, such as cell cycle, metabolism, DNA repair, senescence, and apoptosis. This factor is therefore essential for maintaining genome integrity [1]. In humans, p53 has 393 amino acids and the *TP53* gene is located in the short arm of chromosome 17 [2]. Alterations of the *TP53* gene or perturbations in the *TP53* pathway are frequently correlated with carcinogenesis; more than 50% of human tumors carry mutations in this gene [3].

The steady-state levels of p53 are primarily determined by the rate at which it is degraded, rather than the rate at which it is produced. The *TP53* gene is constitutively expressed in all cell types, but p53 does not accumulate in non-stressed cells, since it is rapidly degraded by the proteasome via ubiquitination [4, 5]. On the other hand, the p53 levels increase in response to various stress signals, such as UV irradiation, low oxygen concentrations (hypoxia), and exposure to high temperatures [6, 7, 8, 9].

There are many polymorphisms described for *TP53*, but a C→G non-synonymous substitution (rs1042522: c.215C>G, p. Pro72Arg; [10]) that promotes the amino acid change Pro→Arg at codon 72 of p53 is one of the most widely studied. This polymorphism has been described to be associated with an increased risk for developing cancer, since the p53-Pro allele is less active than p53-72Arg in inducing apoptosis, among other characteristics [11, 12].

Proper p53 transcriptional function is strongly linked to the activity of several other proteins encoded by the genes *MDM2* (Mouse double minute 2 homolog; OMIM 164785), *MDM4* (Mouse double minute 4 homolog, OMIM 602704), and *USP7* (Ubiquitin-specific protease 7; OMIM 602519), also known as *HAUSP* (Herpesvirus-associated ubiquitin-specific protease). Another important gene in the so-called classical *TP53* network [13] is *LIF* (Leukemia-inhibitory factor; OMIM 159540), which plays an essential role in the early phases of embryonic development in humans, and is regulated by p53 (S1 Fig).

The E3 ubiquitin-protein ligase, MDM2, mediates the activity of p53 by directing it to degradation by the proteasome [5, 14, 15]. *MDM2* expression is also tightly regulated by p53 [16]. This auto-regulatory loop allows for the precise regulation of protein levels and activities of both p53 and MDM2 proteins [4, 17, 18].

The most well studied polymorphism in the *MDM2* gene (rs2279744: c.14+309T>G) is located in its internal promoter. It consists of a single-nucleotide change from T→G, which increases the affinity of a sequence in *MDM2* for the Sp1 transcription factor (Specificity protein 1; OMIM 189906). As a result, homozygotes for the G allele express more MDM2 than homozygotes for the T allele [19, 20]. In the presence of high levels of MDM2, there is a corresponding decrease of p53, causing a reduced response to cellular stress, impaired DNA repair, decreased apoptosis, and senescence [19]. Some studies have demonstrated that *MDM2-309T* and *MDM2-309G* alleles have different distributions in human populations [18, 21]. For instance, derived allele *MDM2-309-G* has higher frequency in European and Asian than African populations (average values: ∼0.35, ∼0.70, and ∼0.03, respectively; [22, 23, 24]). This allele may compensate for the higher apoptotic frequencies caused by the prevalence of allele p53-72Arg in Eurasians (∼0.56; [22, 23, 24]), suggesting adaptation [18].

The MDM4 protein, encoded by the *MDM4* gene acts as a negative regulator of p53, inhibiting its transcriptional activity [25, 26, 27]. MDM2 and MDM4 form heterodimers with a high capacity for ubiquitination of target proteins, thus leading to degradation of targets, like p53 [28]. Deletion of either *MDM2* or *MDM4* induces p53-dependent early embryonic lethality in an animal model [16, 29]. The AA genotype for the single-nucleotide *MDM4* polymorphism (rs1563828:g.204547449A>G) was associated with an increased risk for breast cancer [30].

Another important regulator of p53 is USP7, encoded by the *USP7* gene, which deubiquitylates p53 and protects it from proteasome degradation [31]. The *USP7* gene has a G→A substitution in intron 25 (rs1529916: g.8897333G>A), and derived allele A has been associated with endometriosis, female infertility, and prostate cancer [13, 32].

LIF is a cytokine expressed in various cell types, and its main function is to strengthen the blastocyst training of human embryos. In the very first days post-fertilization, LIF expression increases in the endometrium, creating a favorable environment for blastocyst implantation. Allele *G* of *LIF* (T→G transversion at the 3′ UTR region of the gene; rs929271: g.30242237T>G) is associated with female infertility [13]. LIF expression level is also known to be 2 times lower in cells bearing the p53-72Pro allele, compared to p53-72Arg, which can lead to the decrease of the implantation and fertility rate. In summary, several studies have strongly suggested that polymorphisms in the p53 signaling pathway play an important role in blastocyst implantation and are associated with recurrent pregnancy loss [13, 33, 34].

The genetic variability observed in contemporary human populations and the functionalities associated with the polymorphisms described above allow us to infer that a simple neutral model of mutation and drift is insufficient to explain the allelic distributions observed. Thus, it has been suggested that positive selection contributed to adaptation of *Homo sapiens* in different ecosystems. For example, the p53-72Arg allele (rs1042522) is more common in Europeans than in Africans, leading to the hypothesis that its distribution is dependent on latitude and maintained by selective pressures [35, 36]. On the other hand, Shi *et al.* [23] found that winter temperatures and UV radiation correlated significantly with the *TP53* (rs1042522) and *MDM2* (rs2279744) allele distributions in East Asian populations, indicating the possibility of adaptation to distinct environments.

America was the last continent occupied by humans in pre-colonial times. González-José *et al.* [37] and Bortolini *et al.* [38] suggested that an initial major dispersal began after 21,000 years before present, and that the biological and cultural characteristics of the first Americans that emerged, in part, were reshaped by recurrent trans-Beringian/circum-Arctic gene flow and important local population dynamics during a standstill period in Beringia. For example, Native Americans have experienced dramatic episodes of genetic drift and successive bottleneck events during migration across the continent. Furthermore, signals of positive natural selection associated to autochthonous environmental and cultural conditions have also been described [39, 40, 41].

Based on these findings, we hypothesize that the allele distributions of the classical *TP53* pathway genes in Native American populations reflect adaptation, not only demographic and/or random events. To test our hypothesis, we determined the genotypes of the five above-mentioned SNPs in 282 unrelated individuals and compared the results to a large number of climate-related environmental variables, such as altitude, temperature, and seasonal mean UV radiation. Additional data regarding two of these SNPs (*TP53*-rs1042522 and *MDM2*-rs2279744) were compiled from the literature for a more extensive population analysis.

Materials and Methods

Samples and ethical procedures

Five SNPs (rs929271, rs1042522, rs1563828, rs2279744, and rs1529916) were genotyped in 282 volunteers characterized as Native American or as having large (> 90%; [42]) Native American ancestry. Volunteers were from 12 populations located in different ecoregions, namely highland (populations located at altitudes ≥ 2,500 m; [43]) and lowland (populations located at altitudes below 2,500 m). Highland populations were Aymara (n = 18) and Quechua (n = 17) from Bolivia, and Chivay (n = 18), Cabanaconde (n = 17), Yanke (n = 10), Taquile (n = 43),

Amantani (n = 29), Anapia (n = 15), and Uros (n = 22) from Peru. All highland populations were located in the Andean region, including on Lake Titicaca islands or in their vicinity. Lowland populations were Andoas (n = 61), a Native Amazonian population living in North Peru, and Guaraní Indians from Brazil (Tupian speakers from two sub-groups: Ñandeva, n = 16; and Kaiowa, n = 16). Details about these populations have been summarized elsewhere [42, 44, 45, 46]. To facilitate the presentation of the results and discussion, we will collectively refer to all communities as "Native Americans". The geographical coordinates (latitude and longitude) of all populations are presented in S1 File (Table A in S1 File).

Ethical approval for the use of these samples was obtained from the National Ethics Committee of Brazil (Resolution No. 123/98 CONEP) for individuals from Brazilian tribes; and by the Ethics Committee of Universidad San Martín de Porres, Lima, Peru (Peruvian samples) and Université Paul Sabatier Toulouse, Toulouse, France (Bolivian samples). Written informed consent or verbal informed consent (illiterate persons) was obtained individually from tribal participants. Verbal informed consent was registered in the field, and the institutional review ethics committees approved this procedure. This study was carried out in accordance with the Declaration of Helsinki.

Data from literature

Data from 100 additional individuals from five other Amerindian populations (Surui and Karitiana (Brazil), Piapoco (Colombia), Maya and Pima (México) were included in this study. For more details on these samples, please refer to http://www.cephb.fr/HGDP-CEPH-Panel/[47]. The environmental conditions evaluated for all populations (present study and literature sample) are compiled in S1 File (Table A in S1 File). The environmental data were collected for each population using the SoDa Service and WorldClim (http://www.soda-is.com/[48] and http://www.worldclim.org/[49], respectively; last access: December 19, 2014).

All analyses were performed with two sets of data: (A) 12 South American populations, for which original data regarding five SNPs (rs1042522, rs2279744, rs1529916, rs1563828, and rs929271) were obtained in the present study, and (B) all populations genotyped in this study plus five additional populations, for which *TP53* rs1042522 and *MDM2* rs2279744 data are available in the HGDP-CEPH panel [22].

Laboratory methods

Genomic DNA was obtained from saliva, whole blood, or plasma, using the QIAamp DNA extraction Mini kit (Qiagen; https://www.qiagen.com/br/[50]) according to manufacturer's instructions. Genotyping of the *TP53*-rs1042522, *MDM4*-rs1563828, *USP7*-rs1529916, *LIF*-rs929271, and *MDM2*-rs2279744 SNPs was performed by allelic discrimination using the TaqMan Genotyping Assays (Applied Biosystems; http://www.lifetechnologies.com/br/en/home/brands/applied-biosystems.html [51]). Genotyping of *MDM2* rs2279744 was performed using a customized (assay-by-design) assay using probes `FAM-TCCCGCGCCGCAG` and `VIC-CTCCCGCGCCGAAG`, with primers `5′-CGGGAGTTCAGGGTAAAGGT-3′` (forward) and `5′-ACAGGCACCTGCGATCATC-3′` (reverse).

PCR reactions were carried out in 48-well plates, with each reaction containing: 10 ng of genomic DNA, 2× TaqMan® genotyping Master Mix (Applied Biosystems), specific probes for each SNP (40×), and ultra-pure water for a final reaction volume of 10 μL. The PCR conditions were as follows: 95°C for 10 min, followed by 45 cycles of 95°C for 15 s and 63°C for 60 s. *MDM2* rs2279744 genotyping was also done in 48-well plates, with each reaction containing: 10 ng of genomic DNA, 2× TaqMan® genotyping Master Mix, 5 μM of each primer and probe, and water to reach a final volume of 10 μL. *MDM2* PCR conditions were as follows:

50°C for 2 min, 95°C for 10 min, and 45 cycles of 95°C for 15 s and 60°C for 60 s. All reactions were performed in an Illumina Eco Real-Time PCR System, (http://www.uniscience.com/ [52]) and results were analyzed using an Eco Real-Time PCR System and the Software v5.0 associated with that system. All wet-lab analyses were performed in the Laboratory of Human and Molecular Evolution of the Department of Genetics at Federal University of Rio Grande do Sul in Brazil.

Statistical analyses

Hardy-Weinberg equilibriums were calculated using a web-based program (http://www.oege.org/software/hwe-mr-calc.shtml [53]), and the statistical significance was assessed by Chi-square tests ($p < 0.01$). Analysis of molecular variance (AMOVA) using Arlequin 3.5.1.227 was applied to assess the variance among and within the investigated Native American populations [54, 55, 56].

Allele distributions were tested for possible associations with three groups of environmental conditions: 1) geographic: altitude, latitude, and longitude; 2) annual and seasonal mean UV radiation, and 3) Nineteen climate-related variables (Table A in S1 File). Principal component analysis (PCA) was performed to convert the nineteen possibly correlated bioclimatic variables into a smaller number of artificial variables (PCs) accounting for most of the variance in the previously observed variables. The correlation analysis between allele frequencies in each population and the environmental conditions was performed using Spearman´s rho correlation coefficient. The association between SNPs and altitude was assessed through binary logistic regression using two geographic categories (highlands: $\geq$ 2,500 m, and lowlands: $<$ 2,500 m [43]) as the outcome and SNPs as predictors. Since this analysis was not intended to infer causality relationships, the odds ratio was reported as an estimate of size effect. For these analyses a Bonferroni correction was performed and the alpha was set at 0.01 ($\alpha_{Bonferroni}$ = 0.05/5 SNPs tested). Additionally, we performed the nonparametric Multifactor Dimensionality Reduction (MDR, v3.0.2; [57]) approach to detect potential gene–gene interactions. Thus, we used MDR to incorporate information from our 5 and 2 selected loci (data sets A and B, respectively) and an environmental condition as the outcome (altitude: highland and lowland geographic categories). The percentage of information gain (IG) by each SNP is visualized for each node, while the IG for each pairwise connection between SNPs is visualized for each branch. Thus, the independent main effects of each SNP can be compared to the interaction effect. The *p*-value was calculated based on 10,000 permutations.

Results

Table 1 shows the derived allele frequency for each SNP investigated (individual genotypes can be seen in S2 File). Wide variations were observed in some allele frequencies in both population groups (highland and lowland). For instance, the frequency of MDM2-309-G is about five times higher in Guaraní Ñandeva than Guaraní Kaiowa, which may reflect genetic drift since the split of these two Guaraní partialities occurred less than 2,000 years ago [45]. On the other hand, several highland populations from Peru and Bolivia present similar distributions of *MDM2-309-G*. Most of these highland populations show deviations from the Hardy-Weinberg equilibrium (HWE), especially in Peruvian samples for the *MDM2*-309 locus (Table B in S1 File).

AMOVA analysis, using both data sets (Table 1), indicated that homogeneity and population structures could be seen in both highland and lowland populations. For instance, population structure measured by F_{ST} statistics (*i.e.* the among-populations component of genetic variance) is observed in the two groups considering *TP53* rs1042522 (F_{ST} = 0.068 and 0.054,

Table 1. Derived[1] allele frequencies and AMOVA results.

Population (n)	*TP53* G rs1042522	*MDM2* G rs2279744	*MDM4* G rs1563828	*USP7* A rs1529916	*LIF* G rs929271	Reference
Highlands (≥2.500 m.)						
Amantani (29)	0.86	**0.13**	0.60	0.39	0.77	This study
Anapia (15)	0.77	**0.13**	0.57	0.30	0.47	This study
Cabanaconde (17)	0.88	**0.12**	0.74	0.09	0.32	This study
Chivay (18)	0.83	**0.14**	0.67	0	0.36	This study
Taquile (43)	0.92	**0.05**	0.63	0.23	0.44	This study
Uros (22)	0.93	**0.20**	0.84	0.05	0.27	This study
Yanke (10)	0.65	0.05	0.80	0.10	0.45	This study
Aymara (16–18)[2]	0.78	0.21	0.50	0.14	0.12	This study
Quechua (15–17)[2]	**0.53**	0.21	0.66	0.24	**0.40**	This study
Fst	0.068, $p = 0.017$	-0.020, $p = 0.801$	-0.006, $p = 0.506$	0.068, $p = 0.013$	0.118, $p<0.001$	
Lowlands (<2.500 m.)						
Andoas (61)	0.74	**0.16**	0.53	0.39	0.59	This study
Guarani Kaiowa (16)	0.94	0.07	0.66	0.34	0.56	This study
Guarani Ñandeva (15–16)[2]	0.87	0.33	0.72	0.57	0.63	This study
Karitiana (23–24)[2]	0.62	0.59	ND	ND	ND	Sucheston *et al*. 2011
Maya (23–21)[2]	0.83	0.64	ND	ND	ND	Sucheston *et al*. 2011
Piapoco/Curripaco (12–13)[2]	0.92	0.81	ND	ND	ND	Sucheston *et al*. 2011
Pima (21–24)[2]	0.65	0.69	ND	ND	ND	Sucheston *et al*. 2011
Surui (17–20)[2]	1	**0.32**	ND	ND	ND	Sucheston *et al*. 2011
Fst	0.054, $p = 0.028$	0.274, $p<0.001$	0.029, $p = 0.200$	0.020, $p = 0.243$	-0.044, $p = 1.000$	
Fct	-0.008, $p = 0.516$	0.111, $p = 0.029$	0.001, $p = 0.413$	0.091, $p = 0.043$	0.020, $p = 0.261$	

[1]Defined in comparison with the Chimpanzee sequence. For the loci in **bold** deviations from Hardy-Weinberg Equilibrium were detected.

[2]The number of individuals vary according to the investigated locus. ND: No data available.

doi:10.1371/journal.pone.0137823.t001

for highland and lowland, respectively), while for *MDM2* rs2279744 homogeneity is observed in highland populations ($F_{ST} = -0.020$; $p = 0.801$) while high heterogeneity is observed in lowland populations ($F_{ST} = 0.274$; $p < 0.001$). Only the F_{ST} value observed for *LIF* rs929271 in the highland group (11.8%) is similar to the average estimated across the human genome (12%; [58]). For F_{CT} (between-groups component of variance), the variance is high (11%) for *MDM2* rs2279744 data, indicating a remarkable and significant difference between the allelic distributions of the highland and lowland populations.

Principal component analysis

In data set A, the first principal component (PC1) accounted for 73% of total variance, comprising the following bioclimatic variables: annual mean temperature, mean diurnal range, maximum temperature of warmest month, minimum temperature of coldest month, temperature annual range, mean temperature of wettest quarter, mean temperature of driest quarter,

mean temperature of warmest quarter, mean temperature of coldest quarter, annual precipitation, precipitation of wettest month, precipitation in the driest month, precipitation seasonality, precipitation of wettest quarter, precipitation of driest quarter, precipitation of warmest quarter, and precipitation of coldest quarter. The second principal component (PC2) represented 13% of variance, and comprised temperature seasonality, which is a measure of standard deviation × 100 of average annual daily temperatures.

When we expanded our analysis to data set B, PC1 represented 59% of total variance and comprised sixteen bioclimatic variables: annual mean temperature, mean diurnal range, maximum temperature of warmest month, minimum temperature of coldest month, mean temperature of wettest quarter, mean temperature of driest quarter, mean temperature of warmest quarter, mean temperature of coldest quarter, annual precipitation, precipitation of wettest month, precipitation in the driest month, precipitation seasonality, precipitation of wettest quarter, precipitation of driest quarter, precipitation of warmest quarter, and precipitation of coldest quarter. The second principal component (PC2) represented 23% of variance, and comprised isothermality (the ratio of mean diurnal range to temperature annual range), temperature seasonality, and temperature annual range, all of which are connected with climatic changes by seasonality.

Correlation analyses

Correlation coefficients and their statistical significances are given in Table 2. In data set A, there were significant associations between the G allele of *USP7* (rs1529916) and the annual mean of ultraviolet irradiance (rho = 0.760 $p = 0.004$) and PC1 (rho = −0.741, $p = 0.006$). This allele was also nominally associated with the mean of ultraviolet irradiance in the coldest semester (rho = 0.681 $p = 0.015$) and in the warmest semester (rho = 0.618 $p = 0.032$). *MDM2* (rs2279744) T allele was nominally associated to longitude (rho = −0.587 $p = 0.045$) and the mean of ultraviolet irradiance in the coldest semester (rho = 0.605 $p = 0.037$), while *LIF* (rs929271) T allele was nominally associated to annual mean of ultraviolet irradiance (rho = 0.693 $p = 0.013$) and PC1 (rho = −0.664, $p = 0.018$).

In data set B, there were significant associations between the T allele of *MDM2* (rs2279744) and altitude (rho = 0.673, $p = 0.003$), the mean of ultraviolet irradiance in the coldest semester (rho = 0.827 $p < 0.001$), and PC1 (rho = −0.610 $p = 0.009$). This allele was also nominally associated to PC2 (rho = −0.567, $p = 0.018$).

Binary logistic regression analyses

We performed a binary logistic regression analysis to search for possible associations between SNPs and two geographic categories (Highlands: ≥ 2,500 m; Lowlands: < 2,500 m) using altitude as dependent variable (Table C in S1 File). In data set A, we observed statistically significant associations for *USP7* rs1529916 and *LIF* rs929271 SNPs. Individuals who inhabit the highlands were less likely to carry *USP7*-GA (OR = 0.417, $p = 0.002$) and *USP7*-AA (OR = 0.135, $p < 0.001$) genotypes. A similar association was observed for *LIF*-TG (OR = 0.324, $p = 0.001$ and *LIF*-GG (OR = 0.270, $p < 0.001$) genotypes. Regarding data set B, an association between the *MDM2* rs2279744 SNP and altitude was detected. Individuals who inhabit the highlands were less likely to carry *MDM2*-TG (OR = 0.218, $p < 0.001$) and *MDM2*-GG (OR = 0.175, $p < 0.001$) genotypes.

Gene-gene interaction analyses

We used the MDR approach to search for gene-gene interactions (Table D in S1 File). These analyses were intended to explore differences between highland and lowland populations in

Table 2. Correlation analysis results.

Variables	*TP53* C		*MDM2* T		*MDM4* A		*USP7* G		*LIF* T	
	rs1042522		rs2279744		rs1563828		zrs1529916		rs929271	
	rho	*p*-value	rho	*p*-value	rho	*p*-value	rho	*p*-value	rho	*p*-value
Data set A										
Altitude	0.091	0.778	0.442	0.150	0.330	0.295	0.291	0.359	0.312	0.324
Latitude	0.503	0.095	0.380	0.224	-0.074	0.820	0.301	0.342	0.140	0.665
Longitude	-0.161	0.618	-0.587	*0.045*	0.322	0.307	-0.371	0.236	-0.035	0.914
UV irradiance 1	0.092	0.776	0.542	0.069	-0.143	0.657	0.760	**0.004**	0.693	*0.013*
UV irradiance 2	0.190	0.553	0.605	*0.037*	-0.120	0.710	0.681	*0.015*	0.497	0.100
UV irradiance 3	-0.042	0.897	0.369	0.238	-0.165	0.608	0.618	*0.032*	0.516	0.086
PC1	-0.140	0.665	-0.390	0.210	0.109	0.737	-0.741	**0.006**	-0.664	*0.018*
PC2	-0.140	0.665	-0.464	0.129	-0.175	0.586	-0.049	0.880	0.028	0.931
Data set B										
Altitude	0.006	0.981	0.673	**0.003**	-	-	-	-	-	-
Latitude	0.277	0.281	-0.292	0.255	-	-	-	-	-	-
Longitude	-0.293	0.253	-0.270	0.294	-	-	-	-	-	-
UV irradiance 1	0.265	0.304	0.410	0.102	-	-	-	-	-	-
UV irradiance 2	0.147	0.574	0.827	**<0.001**	-	-	-	-	-	-
UV irradiance 3	0.215	0.408	0.245	0.343	-	-	-	-	-	-
PC1	-0.177	0.497	-0.610	**0.009**	-	-	-	-	-	-
PC2	-0.015	0.955	-0.567	*0.018*	-	-	-	-	-	-

Nominal associations are depicted in *italics* and significant associations (after Bonferroni correction) are depicted in **bold**.

UV irradiance 1: annual mean of ultraviolet irradiance, UV irradiance 2: mean of ultraviolet irradiance in the coldest semester, UV irradiance 3: mean of ultraviolet irradiance in the warmest semester.

PC = Principal Component.

doi:10.1371/journal.pone.0137823.t002

genotype combinations among the SNPs investigated since gene networks, such as those investigated here, can be sources of epistasis. Significant two- ($p = 0.004$) and three-locus interactions ($p = 0.004$) were identified in data set A. However, an analysis of IG based on entropy measures revealed that these effects were not explicated by epistasis (negative values in the branches among nodes; Fig 1A). On the other hand, IG values of both *USP7* (7.26%) and *LIF* (4.23%) indicated that both genes have a large main effect in a scenario where altitude is considered, corroborating our previous analysis. Regarding data set B, the largest main effect was observed for *MDM2* (IG = 9.51%), which contrasts with the low value for *TP53* (IG = 0.41%). A potential synergism (epistasis) between the two loci was also found, but it is apparently weak (IG value of only 1.54%; Fig 1A), at least when it is compared with the potential mechanism of action on *MDM2*. On the other hand, it is 3.7 times greater than the main effect of *TP53*. It is noteworthy that independent of the *TP53* genotype, the genotype *MDM2*-TT is always favorable and most commonly found in highlands (Fig 1B). In other words, *MDM2* showed the greatest contribution to adaptation to hostile environments, such as those found in the highlands.

Discussion

More than 60,000 scientific studies have been published in the last 30 years concerning the roles of *TP53* network genes, as well as of their variants, in human susceptibility to cancer and

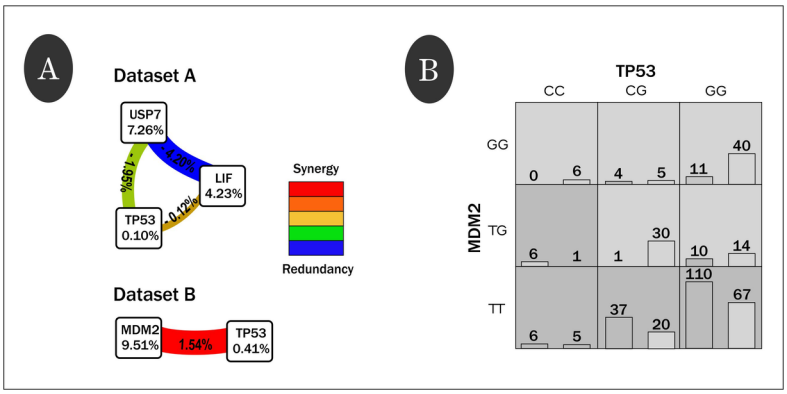


Fig 1. Summary of the multifactor dimensionality reduction (MDR) interaction models. (A) Interaction graphs comprised of nodes with pairwise connections between them. Values in nodes represent information gain (IG) of individual genes (main effect), while values between nodes are the IG of each pairwise combination (interaction effects). Positive entropy (plotted in red) indicates interaction (epistasis) and negative entropy (plotted in green or blue) indicates redundancy. Independence is represented by the gold color. **(B)** The *MDM2*-*TP53* interaction associated with altitude in data set B. High-frequency genotype combinations in individuals who inhabit highlands (≥ 2,500 meters) are depicted as darkly shaded cells and low-frequency combinations in those individuals as lightly shaded. For each cell, the left bar indicates the absolute number of individuals who inhabit highlands and the right bar the absolute number of individuals who inhabit lowlands (< 2,500 meters).

doi:10.1371/journal.pone.0137823.g001

other pathological conditions. Special issues in scientific journals, dedicated to these topics, have also been published (see, as example [59]). This overwhelming number of studies contrasts with the rarity of studies of an evolutionary context, which are indispensable for explaining differences in the *TP53* network allele distributions along human populations, which often cannot be understood as simply a result of stochastic processes. Our goal here was to help fill this gap, providing information about five polymorphisms of the classical *TP53* network in Native American populations and how their variability patterns could be explained.

Our analysis of data set A, which included original information of 5 SNPs in 12 Native American populations, suggests a well-known role of genetic drift in those groups, illustrated by wide difference in *MDM2*-G allele frequencies between the two Guaraní sub-groups. However, other instigating results can be associated to adaptation to environmental conditions in Native American populations. Alleles *USP7*-G (rs1529916) and *LIF*-T (rs929271) were correlated with ultraviolet irradiance and index of temperature and precipitation, variables comprising PC1. Additionally, examining variables with the highest representation in the PC1 components (> 0.90), it is possible to see that in regions where the annual mean temperatures, minimum temperatures of the coldest month, mean temperatures of the driest quarter, mean temperatures of the coldest quarter, and precipitation are low, the presence of ancestral alleles G and T are significantly higher. In other words, our analysis as whole reveals that alleles *USP7*-G and *LIF*-T are more highly represented in stressful environments (low temperature, arid climate, wide temperature range during the day, and high levels of UV radiation), which is typical of high altitudes. It is noteworthy that derived alleles of these SNPs have been associated with cancer susceptibility, infertility, and endometriosis [13, 32], so that the alleles *USP7*-G and *LIF*-T could be considered as protective factors against the consequences of harsh environmental stress.

Human populations living at high altitudes are likely to have developed specific adaptations to support both the harsh conditions described above and low oxygen concentrations (hypoxia; [41]). Monge in 1948 [60] proposed that the hypoxia could reduce fertility in humans.

However, recent studies have shown that the reproductive functioning of populations indigenous to high altitudes is adapted to hypoxia and other extreme conditions [61]. Our results with *USP7* (rs1529916) and *LIF* (rs929271) polymorphisms could be connected with adaptation of the reproductively successful ancestors of modern Andes populations.

In examining data set B, we found the ancestral *MDM2*-T allele is strongly correlated with winter mean UV radiation, altitude, and PC1. The highest representations in the PC1 components (> 0.90) are annual mean temperature, minimum temperature of coldest month, minimum temperature of coldest quarter, and annual precipitation. Allele T is significantly more frequent in communities located at high altitudes experiencing extreme environmental conditions, such as high UV radiation and dry and cold climate. In addition, the binary logistic regression analysis showed that *MDM2*-TT individuals are more frequently found in highlands. *MDM2*-TT homozygotes express typical steady-state levels of MDM2, maintaining an adequate level of p53 [20], and consequently can appropriately respond environmental stresses. An important confounding factor could be admixture with Europeans, which is more important in Andean than in the lowland populations considered here [42, 62, 63]. However, any effect of admixture would be in the opposite direction, since *MDM2*-G frequency is relatively high in Spaniards (0.39; [24]).

The inverse correlation between *MDM2*-T frequencies and winter UV radiation is consistent with the findings of Shi *et al.* [23], which showed that low levels of UV are significantly correlated with genotype *MDM2*-GG in Han Chinese populations, similarly deviating from HWE. These authors suggested that *MDM2*-GG is selected for in areas of low UV activity (at high altitudes, the thinner atmosphere will filter less UV radiation; consequently for every 1000 m increase in altitude, the UV radiation level will increase $\sim$12%; http://www.weather.gov.hk/radiation/tidbit/201012/uv_e.htm [64]). Natural selection can be evoked to explain these results, although the HWE test is considered too weak to detect this phenomenon.

As mentioned above, native Andean populations have successfully adapted to environments with low oxygen concentrations. One gene that contributes to hypoxia adaptation is *EPAS1* (Endothelial PAS domain-containing protein 1, also known as *HIF-2α*, Hypoxia-inducible factor—alpha 2 (OMIM 603349)), which acts by preventing toxicity promoted by hypoxia. This gene plays an important role in both the classical and the expanded *TP53* network. For instance, the alpha subunit of EPAS1 regulates p53 activity, including through prevention of damage-induced degradation and nuclear export of MDM2, stabilizing nuclear p53 [65]. Foll *et al.* [41] confirmed the action of positive selection on *EPAS1* in both Tibetans and Andeans. Furthermore, several studies have revealed a role for p53 and its regulation in physiological and metabolic processes resulting from environments with low oxygen concentrations [8, 66, 67]. Recently, Eichstaedt *et al.* [68] studied an indigenous population living in the Argentinean Andes (Colla) and identified signatures of positive selection in genes involved in cellular hypoxia, including *TP53*. Importantly, hypoxia induces p53 accumulation through down-regulation of MDM2 [66]. These results reinforce our suggestion that individuals with the *MDM2*-TT genotype represent an adaptation to the environmental stresses of high altitudes. In addition, the interaction analysis performed by the MDR method using both data sets (A and B) revealed the potential for the *MDM2*, *LIF*, and *USP7* genes to play an additional central role in a high altitude setting. Thus, taken together, our results demonstrate that variation of the p53-activating stressors could not be directly correlated with p53-Pro72Arg alleles, but with frequencies of the other functional polymorphisms examined, such as *USP7*-G (rs1529916), *LIF*-T (rs929271), and *MDM2*-309, as well as synergic interactions between them.

Under neutral model conditions, South Amerindians living in lowlands present higher levels of population structure when compared to those seen in indigenous Andean communities [62, 69]. However, not all F_{ST} values obtained in our study were consistent with this

expectation (Table 1). Positive selection disturbs the patterns of genetic variation expected under a standard neutral model [70]. Additionally, it is possible to see that some derived alleles, such as *MDM2*-G, have high frequencies in Asian populations with putative common ancestry (0.57–0.82; [22, 23, 24]), but a surprisingly low distribution in Andeans (average value: ∼0.13). An excess of unexpectedly low and/or high frequencies of derived alleles can also be considered a marker of positive selection [70]. Thus, the distributions of the classical *TP53* pathway alleles in Native American populations could be under selective pressure. Sucheston *et al.* [22] investigated 52 worldwide populations from the HGDP-CEPH-panel for the prevalence of p53-Pro72Arg and *MDM2*-309 polymorphisms, but found no significant association with climate variables. However, the Native American samples in the Sucheston *et al.*' study [22] were much smaller than the present study (see Table 1), which may explain the divergent results.

Finally, government surveys in Peru indicate that the rate of gestational and postpartum complications in Aymara regions is lower than the national average (1.8% and 5% respectively; http://www.dge.gob.pe/publicaciones/pub_asis/asis26.pdf, p. 165; [71] http://www.dge.gob.pe/portal/docs/intsan/asis2012.pdf, p. 76 [72]). These same official sources also indicate differences in the cancer incidences between lowland localities and some regions situated at high altitude (for example in the Puno state, where the Anapia community is located; http://www.dge.gob.pe/portal/docs/asis_cancer.pdf, p. 64 [73]). These findings are in agreement with our genetic results. However, only additional and specific studies can accurately relate our evolutionary findings with those related to the health of contemporary Andean populations.

A well-regulated p53 network is crucial for maintaining genomic integrity. Several polymorphisms in this pathway have been described, and the different allele frequencies among human populations have been interpreted as the result of selective pressure. Humans occupied high-altitude locations in the Andes as early as 12,800 years ago, providing a sufficient period of time for the initiation of organismal selection and developmental functional adaptation ([74] and references therein). Here we are suggesting that natives from Andes, who are subjected to low temperatures, arid climates, wide temperature ranges during the day, high levels of UV radiation, and hypoxia, among other environmental insults, are protected by a selected combination of alleles/genotypes of the *TP53* pathway. The present study identifies for the first time the potential role of the *MDM2*, *LIF*, and *USP7* in the adaptation of the Andean populations.

Supporting Information

S1 Fig. The p53 network. Network view of p53 pathway analyzed by STRING 10.0 (http://string-db.org/). Interaction confidence score cutoff was 900 (highest confidence). Each color arrow represents a predicted functional partner: green (activation), red (inhibition), blue (binding), purple (catalysis), pink (post-translational modification), black (reaction), and yellow (expression). TP53 = tumor protein p53, USP7 = ubiquitin specific peptidase 7 (herpes virus-associated), MDM4 = Mouse double minute 4 homolog, MDM2 = Mouse double minute 2 homolog, and LIF = leukemia inhibitory factor.
(TIFF)

S1 File. Additional Results. Climatic variables evaluated in population of this study **(Table A)**. Allelic frequencies and Hardy-Weinberg Equilibrium results **(Table B)**. Binary logistic regression analyses results **(Table C)**. Locus interaction by the multifactor dimensionality reduction (MDR) approach **(Table D)**.
(DOC)

S2 File. Individual Genotype Database. In this database, we found individual genotypes of all polymorphisms discussed in the study.
(XLSX)

Acknowledgments

We are very grateful to the individuals who donated the samples analyzed here and to the Fundação Nacional do Indio (Brazil) for logistic support. We thank René Vasquez for his assistance in the sample collection in Bolivia and Sidia Maria Callegari Jacques for statistical support. We are grateful to professor David Comas for his careful review of this manuscript.

Author Contributions

Conceived and designed the experiments: VCJ VR MCB. Performed the experiments: VCJ VR. Analyzed the data: VCJ DLR OP VR. Contributed reagents/materials/analysis tools: SA GSM JRS AS-G MV J-MD RB-M MLP-E FMS PA-P. Wrote the paper: VCJ DLR VR FMS MCB.

References

1. Botcheva K. p53 binding to human genome: crowd control navigation in chromatin context. Front Genet. 2014 Dec 22; 5:447. doi: 10.3389/fgene.2014.00447 PMID: 25566329
2. Linzer DI, Levine AJ. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. Cell. 1979 May; 17(1):43–52. PMID: 222475
3. Leroy B, Anderson M, Soussi T. TP53 mutations in human cancer: database reassessment and prospects for the next decade. Hum Mutat. 2014 Jun; 35(6):672–88. doi: 10.1002/humu.22552 PMID: 24665023
4. Nag S, Qin J, Srivenugopal KS, Wang M, Zhang R. The MDM2-p53 pathway revisited. J Biomed Res. 2013 Jul; 27(4):254–71. doi: 10.7555/JBR.27.20130030 PMID: 23885265
5. Chao CC. Mechanisms of p53 degradation. Clin Chim Acta. 2015 Jan 1; 438:139–47. doi: 10.1016/j.cca.2014.08.015 PMID: 25172038
6. Ljungman M. Dial 9-1-1 for p53: mechanisms of p53 activation by cellular stress. Neoplasia. 2000 May-Jun; 2(3):208–25. PMID: 10935507
7. Latonen L, Taya Y, Laiho M. UV-radiation induces dose-dependent regulation of p53 response and modulates p53-HDM2 interaction in human fibroblasts. Oncogene. 2001 Oct 11; 20(46):6784–93. PMID: 11709713
8. Sermeus A, Michiels C. Reciprocal influence of the p53 and the hypoxic pathways. Cell Death Dis. 2011 May 26; 2:e164. doi: 10.1038/cddis.2011.48 PMID: 21614094
9. Chen J. The Roles of MDM2 and MDMX Phosphorylation in Stress Signaling to p53. Genes Cancer. 2012 Mar; 3(3–4):274–82. doi: 10.1177/1947601912454733 PMID: 23150760
10. Matlashewski GJ, Tuck S, Pim D, Lamb P, Schneider J, Crawford LV. Primary structure polymorphism at amino acid residue 72 of human p53. Mol Cell Biol. 1987 Feb; 7(2):961–3. PMID: 3547088
11. Dumont P, Leu JI, Della Pietra AC 3rd, George DL, Murphy M. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet. 2003 Mar; 33(3):357–65. PMID: 12567188
12. Whibley C, Pharoah PD, Hollstein M. p53 polymorphisms: cancer implications. Nat Rev Cancer. 2009 Feb; 9(2):95–107. doi: 10.1038/nrc2584 PMID: 19165225
13. Kang HJ, Feng Z, Sun Y, Atwal G, Murphy ME, Rebbeck TR, et al. Single-nucleotide polymorphisms in the p53 pathway regulate fertility in humans. Proc Natl Acad Sci U S A. 2009 Jun 16; 106(24):9761–6. doi: 10.1073/pnas.0904280106 PMID: 19470478
14. Momand J, Zambetti GP, Olson DC, George D, Levine AJ. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell. 1992 Jun 26; 69(7):1237–45. PMID: 1535557
15. Marine JC, Francoz S, Maetens M, Wahl G, Toledo F, Lozano G. Keeping p53 in check: essential and synergistic functions of Mdm2 and Mdm4. Cell Death Differ.2006 Jun; 13(6):927–34. PMID: 16543935
16. Eischen CM, Lozano G. The Mdm network and its regulation of p53 activities: a rheostat of cancer risk. Hum Mutat. 2014 Jun; 35(6):728–37. doi: 10.1002/humu.22524 PMID: 24488925

17. Kohn KW, Pommier Y. Molecular interaction map of the p53 and Mdm2 logic elements, which control the Off-On switch of p53 in response to DNA damage. Biochem Biophys Res Commun. 2005 Jun 10; 331(3):816–27. PMID: 15865937

18. Atwal GS, Bond GL, Metsuyanim S, Papa M, Friedman E, Distelman-Menachem T, et al. Haplotype structure and selection of the MDM2 oncogene in humans. Proc Natl Acad Sci U S A. 2007 Mar 13; 104 (11):4524–9. PMID: 17360557

19. Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC, et al. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. Cell. 2004 Nov 24; 119(5):591–602. PMID: 15550242

20. Alazzouzi H, Suriano G, Guerra A, Plaja A, Espín E, Armengol M, et al. Tumour selection advantage of non-dominant negative P53 mutations in homozygotic MDM2-SNP309 colorectal cancer cells. J Med Genet. 2007 Jan; 44(1):75–80. Epub 2006 Jul 6. PMID: 16825434

21. Millikan RC, Heard K, Winkel S, Hill EJ, Heard K, Massa B, et al. No association between the MDM2–309 T/G promoter polymorphism and breast cancer in African-Americans or Whites. Cancer Epidemiol Biomarkers Prev. 2006 Jan; 15(1):175–7. PMID: 16434608

22. Sucheston L, Witonsky DB, Hastings D, Yildiz O, Clark VJ, Di Rienzo A, et al. Natural selection and functional genetic variation in the p53 pathway. Hum Mol Genet. 2011 Apr 15; 20(8):1502–8. doi: 10.1093/hmg/ddr028 PMID: 21266458

23. Shi H, Tan SJ, Zhong H, Hu W, Levine A, Xiao CJ, et al. Winter temperature and UV are tightly linked to genetic changes in the p53 tumor suppressor pathway in Eastern Asia. Am J Hum Genet. 2009 Apr; 84 (4):534–41. doi: 10.1016/j.ajhg.2009.03.009 PMID: 19344876

24. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1; 491 (7422):56–65. doi: 10.1038/nature11632 PMID: 23128226

25. Shvarts A, Steegenga WT, Riteco N, van Laar T, Dekker P, Bazuine M, et al. MDMX: a novel p53-binding protein with some functional properties of MDM2. EMBO J. 1996 Oct 1; 15(19):5349–57. PMID: 8895579

26. Shvarts A, Bazuine M, Dekker P, Ramos YF, Steegenga WT, Merckx G, et al. Isolation and identification of the human homolog of a new p53-binding protein, Mdmx. Genomics. 1997 Jul 1; 43(1):34–42. PMID: 9226370

27. Liu J, Tang X, Li M, Lu C, Shi J, Zhou L, et al. Functional MDM4 rs4245739 genetic variant, alone and in combination with P53 Arg72Pro polymorphism, contributes to breast cancer susceptibility. Breast Cancer Res Treat. 2013 Jul; 140(1):151–7. doi: 10.1007/s10549-013-2615-x PMID: 23793604

28. Pei D, Zhang Y, Zheng J. Regulation of p53: a collaboration between Mdm2 and Mdmx. Oncotarget. 2012 Mar; 3(3):228–35. PMID: 22410433

29. Xiong S. Mouse models of Mdm2 and Mdm4 and their clinical implications. Chin J Cancer. 2013 Jul; 32 (7):371–5. doi: 10.5732/cjc.012.10286 PMID: 23327795

30. Song CG, Fu FM, Wu XY, Wang C, Shao ZM. Correlation of polymorphism rs1563828 in MDM4 gene with breast cancer risk and onset age. Zhonghua Wai Ke Za Zhi. 2012 Jan 1; 50(1):53–6. PMID: 22490292

31. Shan J, Brooks C, Kon N, Li M, Gu W. Dissecting roles of ubiquitination in the p53 pathway. Ernst Schering Found Symp Proc. 2008;(1:):127–36. PMID: 19202598

32. Sun T, Lee GS, Oh WK, Pomerantz M, Yang M, Xie W, et al. Single-nucleotide polymorphisms in p53 pathway and aggressiveness of prostate cancer in a Caucasian population. Clin Cancer Res. 2010 Nov 1; 16(21):5244–51. doi: 10.1158/1078-0432.CCR-10-1261 PMID: 20855462

33. Dd Paskulin, Paixão-Côrtes VR, Hainaut P, Bortolini MC, Ashton-Prolla P. The TP53 fertility network. Genet Mol Biol. 2012 Dec; 35(4 (suppl)):939–46.

34. Fraga LR, Dutra CG, Boquett JA, Vianna FS, Gonçalves RO, Paskulin DD, et al. p53 signaling pathway polymorphisms associated to recurrent pregnancy loss. Mol Biol Rep. 2014 Mar; 41(3):1871–7. doi: 10.1007/s11033-014-3036-6 PMID: 24435975

35. Beckman G, Birgander R, Själander A, Saha N, Holmberg PA, Kivelä A, et al. Is p53 polymorphism maintained by natural selection? Hum Hered. 1994; 44:266–270. PMID: 7927355

36. Själander A, Birgander R, Saha N, Beckman L, Beckman G. p53 polymorphisms and haplotypes show distinct differences between major ethnic groups. Hum Hered. 1996 Jan-Feb; 46(1):41–8. PMID: 8825462

37. González-José R, Bortolini MC, Santos FR, Bonatto SL. The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. Am J Phys Anthropol. 2008 Oct; 137(2):175–87. doi: 10.1002/ajpa.20854 PMID: 18481303

38. Bortolini MC, González-José R, Bonatto SL, Santos FR. Reconciling pre-Columbian settlement hypotheses requires integrative, multidisciplinary, and model-bound approaches. Proc Natl Acad Sci U S A. 2014 Jan 14; 111(2):E213–4. doi: 10.1073/pnas.1321197111 PMID: 24398530

39. Hünemeier T, Amorim CE, Azevedo S, Contini V, Acuña-Alonzo V, Rothhammer F, et al. Evolutionary responses to a constructed niche: ancient Mesoamericans as a model of gene-culture coevolution. PLoS One. 2012; 7(6):e38862. doi: 10.1371/journal.pone.0038862 PMID: 22768049

40. Hünemeier T, Gómez-Valdés J, Ballesteros-Romero M, de Azevedo S, Martínez-Abadías N, Esparza M, et al. Cultural diversification promotes rapid phenotypic evolution in Xavánte Indians. Proc Natl Acad Sci U S A. 2012 Jan 3; 109(1):73–7. doi: 10.1073/pnas.1118967109 PMID: 22184238

41. Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L. Widespread signals of convergent adaptation to high altitude in Asia and america. Am J Hum Genet. 2014 Oct 2; 95(4):394–407. doi: 10.1016/j.ajhg.2014.09.002 PMID: 25262650

42. Sandoval JR, Salazar-Granara A, Acosta O, Castillo-Herrera W, Fujita R, Pena SD, et al. Tracing the genomic ancestry of Peruvians reveals a major legacy of pre-Columbian ancestors. J Hum Genet. 2013 Sep; 58(9):627–34. doi: 10.1038/jhg.2013.73 PMID: 23863748

43. Moore LG. Human genetic adaptation to high altitude. High Alt Med Biol. 2001 Summer; 2(2):257–79. PMID: 11443005

44. Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfin LA, Zago MA, et al. HLA class II diversity in seven Amerindian populations. Clues about the origins of the Aché. Tissue Antigens. 2003 Dec; 62(6):512–26. PMID: 14617035

45. Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, et al. Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. Am J Phys Anthropol. 2007 Feb; 132(2):301–10. PMID: 17133437

46. Gayà-Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodríguez A, Villena M, et al. Autosomal and X chromosome *Alu* insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. Am J Hum Biol. 2010 Mar-Apr; 22(2):154–62. doi: 10.1002/ajhb.20967 PMID: 19593738

47. Available: http://www.cephb.fr/HGDP-CEPH-Panel/. Accessed October 2014.

48. http://www.soda-is.com/. Accessed 19 December 19 2014.

49. http://www.worldclim.org/. Accessed 19 December 2014.

50. https://www.qiagen.com/br/

51. www.lifetechnologies.com/br/en/home/brands/applied-biosystems.html

52. http://www.uniscience.com/

53. http://www.oege.org/software/hwe-mr-calc.shtml,, last accessed January 2015.

54. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 1984; 38: 1358–1370.

55. Weir BS. The second National Research Council report on forensic DNA evidence.bAm J Hum Genet. 1996 Sep; 59(3):497–500.

56. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992 Jun; 131(2):479–91. PMID: 1644282

57. Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, et al. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol. 2006 Jul 21; 241(2):252–61. PMID: 16457852

58. International HapMap Consortium. A haplotype map of the human genome. Nature. 2005 Oct 27; 437 (7063):1299–320. PMID: 16255080

59. Soussi T. The TP53 gene network in a postgenomic era. Hum Mutat. 2014 Jun; 35(6):641–2. PMID: 24753184

60. Monge C. Acclimatization in the Andes. Baltimore: Johns Hopkins. University Press; 1948.

61. Vitzthum VJ. Fifty fertile years: anthropologists' studies of reproduction in high altitude natives. Am J Hum Biol. 2013 Mar-Apr; 25(2):179–89. doi: 10.1002/ajhb.22357 PMID: 23382088

62. Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, et al. Genetic variation and population structure in native Americans. PLoS Genet. 2007 Nov; 3(11):e185. PMID: 18039031

63. Verdu P, Pemberton TJ, Laurent R, Kemp BM, Gonzalez-Oliver A, Gorodezky C, et al. Patterns of admixture and population structure in native populations of Northwest North America. PLoS Genet. 2014 Aug 14; 10(8):e1004530. doi: 10.1371/journal.pgen.1004530 PMID: 25122539

64. http://www.weather.gov.hk/radiation/tidbit/201012/uv_e.htm. Accessed January 2015.

65. Chen D, Li M, Luo J, Gu W. Direct interactions between HIF-1 alpha and Mdm2 modulate p53 function. J Biol Chem. 2003 Apr 18; 278(16):13595–8. PMID: 12606552

66. Alarcón R, Koumenis C, Geyer RK, Maki CG, Giaccia AJ. Hypoxia induces p53 accumulation through MDM2 down-regulation and inhibition of E6-mediated degradation. Cancer Res. 1999 Dec 15; 59 (24):6046–51. PMID: 10626788

67. Vousden KH, Ryan KM. p53 and metabolism. Nat Rev Cancer. 2009; 9:691–700. doi: 10.1038/nrc2715 PMID: 19759539

68. Eichstaedt CA, Antão T, Pagani L, Cardona A, Kivisild T, Mormina M. The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS One. 2014 Mar 31; 9(3):e93314. doi: 10.1371/journal.pone.0093314 PMID: 24686296

69. Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, et al. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet. 2001 Jun; 68(6):1485–96. PMID: 11353402

70. Biswas S, Akey JM. Genomic insights into positive selection. Trends Genet. 2006; 22:437–446. PMID: 16808986

71. http://www.dge.gob.pe/publicaciones/pub_asis/asis26.pdf, p. 165;

72. http://www.dge.gob.pe/portal/docs/intsan/asis2012.pdf, p. 76;

73. http://www.dge.gob.pe/portal/docs/asis_cancer.pdf, p. 64;

74. Rademaker K, Hodgins G, Moore K, Zarrillo S, Miller C, Bromley GR, et al. Paleoindian settlement of the high-altitude Peruvian Andes. Science. 2014 Oct 24; 346(6208):466–9. doi: 10.1126/science.1258260 PMID: 25342802

S1 File. Additional Results.

Table A. Climatic variables evaluated in population of this study.

Populations	Aymara	Quechua	G. Ñandeva	G. Kaiowá	Amantani	Andoas	Anapia	Cabanaconde	Chivay	Taquile	Uros	Yanke	Karitiana	Maya	Pima	Surui	Piapoco/ Curripaco
Altitude	4312	2656	319	456	3827	235	4366	3441	3820	3825	3818	3820	86	3	361	303	117
Latitude	-19.25	-14.5	-23.8	-23.1	-15.65	-2.9	-16.31	-15.62	-15.63	-15.76	-15.74	-15.64	-8.75	20.22	33.17	-8.4	3.0
Longitude	-69.08	-69.0	-54.5	-55.2	-69.71	-76.4	-68.85	-71.98	-71.98	-69.68	-69.93	-71.95	-63.84	-90.47	-11.87	-54.9	-68.0
UV irradiance 1	230	202	212	210	212	206	212	248	224	216	214	226	180	216	246	180	162
UV irradiance 2	210	198	150	154	206	209	204	241	214	206	206	214	194	184	183	176	172
UV irradiance 3	252	205	212	210	220	203	219	219	235	220	221	238	165	246	309	161	153
Bio 1	32	166	218	221	83	257	86	132	83	84	85	93	260	267	212	253	267
Bio 2	181	113	117	118	114	99	137	171	183	115	131	182	103	102	179	135	104
Bio 3	71	72	57	61	70	86	65	79	73	71	71	74	71	62	46	69	83
Bio 4	2664	1112	2748	2595	1245	495	1920	1070	1632	1216	1370	1540	631	2060	7809	523	586
Bio 5	145	236	314	309	155	316	174	232	191	156	167	200	329	347	410	354	337
Bio 6	-109	81	112	116	-7	202	-36	16	-57	-4	-16	-43	185	183	24	161	212
Bio 7	254	155	202	193	162	114	210	216	248	160	183	243	144	164	386	248	260
Bio 8	61	173	237	238	93	253	104	142	99	94	96	109	259	282	308	248	260
Bio 9	16	148	186	186	64	262	58	119	64	66	64	70	258	264	247	251	272
Bio 10	61	177	249	250	94	262	105	142	99	95	98	109	268	288	313	259	274
Bio 11	-7	148	179	185	64	250	58	116	58	66	64	70	252	237	114	247	260
Bio 12	194	1668	1593	1566	964	2552	728	289	453	1134	743	430	2124	867	206	2415	3192
Bio 13	79	276	194	191	210	260	161	75	107	249	158	106	310	177	33	385	459
Bio 14	0	25	66	57	6	165	7	1	1	7	3	1	21	11	3	9	95
Bio 15	159	62	29	33	84	13	81	107	105	83	89	107	61	79	48	69	46
Bio 16	179	773	553	555	539	728	400	205	307	632	435	295	921	455	77	1083	1266
Bio 17	1	101	246	213	27	573	28	7	9	34	16	8	100	36	13	49	363
Bio 18	179	537	482	498	420	573	333	205	268	298	318	260	319	199	61	343	377
Bio 19	1	101	322	266	27	646	28	9	9	34	16	8	353	62	62	1082	1233

UV irradiance 1: annual mean of ultraviolet irradiance , UV irradiance 2: mean of ultraviolet irradiance in the coldest semester, UV irradiance 3: mean of ultraviolet irradiance in the warmest semester; Bio = Bioclimatic variables, Bio 1: Annual Mean Temperature, Bio 2: Mean Diurnal Range, Bio 3: Isothermality, Bio 4: Temperature Seasonality, Bio 5: Maximum Temperature of Warmest Month, Bio 6: Minimum Temperature of Coldest Month, Bio 7: Temperature Annual Range, Bio 8: Mean Temperature of Wettest Quarter, Bio 9: Mean Temperature of Driest Quarter, Bio 10: Mean Temperature of Warmest Quarter, Bio 11: Mean Temperature of Coldest Quarter, Bio 12: Annual Precipitation, Bio 13: Precipitation of Wettest Month, Bio 14: Precipitation in the driest month, Bio 15: Precipitation Seasonality, Bio 16: Precipitation of Wettest Quarter, Bio 17: Precipitation of Driest Quarter, Bio 18: Precipitation of Warmest Quarter, Bio 19: Precipitation of Coldest Quarter. All temperature indexes are in °C × 10 and precipitation indexes are in mm. Irradiance is in W/m².

S1 File. Additional Results.

Table B. Allelic frequencies and Hardy-Weinberg Equilibrium results.

Population	Country	*TP53* rs1042522			*MDM2* rs2279744			*MDM4* rs1563828			*USP7* rs1529916			*LIF*rs929271			Reference
		C	G	n	T	G	n	A	G	n	G	A	n	T	G	n	
Highlands(> 2.500 m.)																	
Amantani	Peru	0.14	0.86	29	**0.87**	**0.13**	29	0.40	0.60	29	0.61	0.39	29	0.23	0.77	29	This study
Anapia	Peru	0.23	0.77	15	**0.87**	**0.13**	15	0.43	0.57	15	0.70	0.30	15	0.53	0.47	15	This study
Cabanaconde	Peru	0.12	0.88	17	**0.88**	**0.12**	17	0.26	0.74	17	0.91	0.09	17	0.68	0.32	17	This study
Chivay	Peru	0.17	0.83	18	**0.86**	**0.14**	18	0.33	0.67	18	1	0	18	0.64	0.36	18	This study
Taquile	Peru	0.08	0.92	43	**0.95**	**0.05**	43	0.37	0.63	43	0.77	0.23	43	0.56	0.44	43	This study
Uros	Peru	0.07	0.93	22	**0.80**	**0.20**	22	0.16	0.84	22	0.95	0.05	22	0.73	0.27	22	This study
Yanke	Peru	0.35	0.65	10	0.95	0.05	10	0.20	0.80	10	0.90	0.10	10	0.55	0.45	10	This study
Aymara	Bolivia	0.22	0.78	16	0.79	0.21	17	0.50	0.50	16	0.86	0.14	18	0.88	0.12	17	This study
Quechua	Bolivia	**0.47**	**0.53**	16	0.79	0.21	17	0.34	0.66	16	0.76	0.24	17	**0.60**	**0.40**	15	This study
Lowlands (< 2.500 m.)																	
Andoas	Peru	0.26	0.74	61	**0.84**	**0.16**	61	0.47	0.53	61	0.61	0.39	61	0.41	0.59	61	This study
Guarani Kaiowa	Brazil	0.06	0.94	16	0.93	0.07	16	0.34	0.66	16	0.66	0.34	16	0.44	0.56	16	This study
Guarani Ñandeva	Brazil	0.13	0.87	15	0.67	0.33	15	0.28	0.72	16	0.43	0.57	15	0.37	0.63	15	This study
Karitiana	Brazil	0.38	0.62	24	0.41	0.59	23	ND	ND	ND	ND	ND	ND	ND	ND	ND	Sucheston *et al.* 2011
Maya	Mexico	0.17	0.83	23	0.36	0.64	21	ND	ND	ND	ND	ND	ND	ND	ND	ND	Sucheston *et al.* 2011
Piapoco/Curripaco	Colombia	0.08	0.92	12	0.19	0.81	13	ND	ND	ND	ND	ND	ND	ND	ND	ND	Sucheston *et al.* 2011
Pima	USA	0.35	0.65	24	0.31	0.69	21	ND	ND	ND	ND	ND	ND	ND	ND	ND	Sucheston *et al.* 2011
Surui	Brazil	0	1	17	**0.68**	**0.32**	20	ND	ND	ND	ND	ND	ND	ND	ND	ND	Sucheston *et al.* 2011

For the loci **in bold** deviations from the Hardy-Weinberg Equilibrium were detected. ND: No data available.

S1 File. Additional Results.

Table C. Binary logistic regression analyses results.

	B	SE	Wald	df	OR	CI 95 %	*p*-value
Data set A							
TP53 rs1042522							
GG (reference)	0				1		
GC	-0.212	0.299	0.504	1	0.809	0.450-1.453	0.478
CC	-0.233	0.500	0.217	1	0.792	0.297-2.111	0.641
MDM2 rs2279744							
TT (reference)	0				1		
TG	-0.179	0.419	0.182	1	0.836	0.368-1.902	0.670
GG	-0.361	0.432	0.698	1	0.697	0.299-1.626	0.403
MDM4 rs1563828							
GG (reference)	0				1		
GA	-0.444	0.281	2.489	1	0.642	0.370-1.114	0.115
AA	-0.552	0.375	2.172	1	0.576	0.276-1.200	0.141
USP7 rs1529916							
GG (reference)	0				1		
GA	-0.875	0.285	9.415	1	0.417	0.238-0.729	**0.002**
AA	-2.000	0.447	20.054	1	0.135	0.056-0.325	**<0.001**
LIF rs929271							
TT (reference)	0				1		
TG	-1.126	0.342	10.875	1	0.324	0.166-0.633	**0.001**
GG	-1.310	0.368	12.681	1	0.270	0.131-0.555	**<0.001**
Data set B							
TP53 rs1042522							
GG (reference)	0				1		
GC	-0.427	0.238	3.220	1	0.653	0.409-1.040	0.073
CC	-0.087	0.427	0.041	1	0.917	0.397-2.118	0.839
MDM2 rs2279744							
TT (reference)	0				1		
TG	-1.523	0.305	24.868	1	0.218	0.120-0.397	**<0.001**
GG	-1.745	0.322	29.407	1	0.175	0.093-0.328	**<0.001**

SE = standard error; df = degrees of freedom; OR = odds ratio; CI = confidence interval.

Table D. Locus interaction by the multifactor dimensionality reduction (MDR) approach.

Sample	Model	Testing accuracy	*p*-value*	Cross-validation consistency
Data set A				
	USP7	0.6438	0.004	10/10
	LIF,USP7	0.6409	0.004	10/10
	LIF,TP53,USP7	0.6408	0.004	9/10
Data set B				
	MDM2	0.6688	0.001	10/10
	TP53, MDM2	0.6690	0.001	10/10

*Evaluated using a 1000-fold permutation test to compare observed testing accuracies with those expected under the null hypothesis of null association.

S2 File. Individual Genotype Database

LIF

rs929271
chromosome 22
Chr position 28968226

Aymaras	TT	GG	GT
BOA 158	1		
BOA 161	1		
BOA 174	1		
BOA 175	1		
BOA 185	1		
BOA 204	1		
BOA 239	1		
BOA 241			1
BOA 243	1		
BOA 250			1
BOA 259	1		
BOA 267	1		
BOA 269	1		
BOA 271			1
BOA 277			1
BOA 292	1		
BOA 314	1		
Total	13	0	4

Quechuas	TT	GG	TG
BOQ 516		1	
BOQ 523	1		
BOQ 529	1		
BOQ 537			1
BOQ 555		1	
BOQ 586	1		
BOQ 602		1	
BOQ 607	1		
BOQ 624	1		
BOQ 639	1		
BOQ 643		1	
BOQ 672	1		
BOQ 693			1
BOQ 702	1		
BOQ 723		1	
Total	8	5	2

Ñandeva	TT	GG	TG
G. Ñandeva 5		1	
G. Ñandeva 6		1	
G. Ñandeva 20		1	
G. Ñandeva 22		1	
G. Ñandeva 24			1
G. Ñandeva 28			1
G. Ñandeva 29			1
G. Ñandeva 405			1
G. Ñandeva 825	1		
G. Ñandeva 829			1
G. Ñandeva 835			1
G. Ñandeva 837			1
G. Ñandeva 841			1
G. Ñandeva 844			1
G. Ñandeva 852		1	
Total	1	5	9

Kaiowa	TT	GG	TG
G. Kaiowa 203	1		
G. Kaiowa 205			1
G. Kaiowa 220		1	
G. Kaiowa 223	1		
G. Kaiowa 224			1
G. Kaiowa 229			1
G. Kaiowa 230	1		
G. Kaiowa 603		1	
G. Kaiowa 608			1
G. Kaiowa 626		1	
G. Kaiowa 634		1	
G. Kaiowa 644		1	

TP53

rs1042522
chromosome 17
Chr position 7520197

Aymaras	GG	CC	CG
BOA 158	1		
BOA 161			1
BOA 174			1
BOA 175	1		
BOA 185			1
BOA 204	1		
BOA 239		1	
BOA 241	1		
BOA 243	1		
BOA 250	1		
BOA 259	1		
BOA 267	1		
BOA 269	1		
BOA 271	1		
BOA 277	1		
BOA 292		1	
Total	11	2	3

Frec G 0,78

Quechuas	GG	CC	CG
BOQ 513	1		
BOQ 516	1		
BOQ 523		1	
BOQ 529	1		
BOQ 555		1	
BOQ 586		1	
BOQ 602	1		
BOQ 607		1	
BOQ 624	1		
BOQ 639	1		
BOQ 643	1		
BOQ 672		1	
BOQ 686	1		
BOQ 693			1
BOQ 702		1	
BOQ 723		1	
Total	8	7	1

Ñandeva	GG	CC	CG
G. Ñandeva 5	1		
G. Ñandeva 6	1		
G. Ñandeva 20	1		
G. Ñandeva 22	1		
G. Ñandeva 24	1		
G. Ñandeva 28	1		
G. Ñandeva 405	1		
G. Ñandeva 825	1		
G. Ñandeva 829	1		
G. Ñandeva 835	1		
G. Ñandeva 837	1		
G. Ñandeva 841			1
G. Ñandeva 844			1
G. Ñandeva 845		1	
G. Ñandeva 852	1		
Total	12	1	2

Kaiowa	GG	CC	CG
G. Kaiowa 203	1		
G. Kaiowa 205	1		
G. Kaiowa 220	1		
G. Kaiowa 223	1		
G. Kaiowa 224	1		
G. Kaiowa 229	1		
G. Kaiowa 230	1		
G. Kaiowa 603			1
G. Kaiowa 608	1		
G. Kaiowa 626	1		
G. Kaiowa 634	1		
G. Kaiowa 644			1

MDM4

rs1563828
chromosome 1
Chr position: 202783200

Aymaras	AA	GG	AG
BOA 158		1	
BOA 161			1
BOA 169	1		
BOA 174		1	
BOA 175		1	
BOA 185			1
BOA 204			1
BOA 239	1		
BOA 241	1		
BOA 243	1		
BOA 259		1	
BOA 267			1
BOA 269	1		
BOA 271			1
BOA 277			1
BOA 292		1	
Total	5	5	6

Quechuas	AA	GG	AG
BOQ 516			1
BOQ 523			1
BOQ 529	1		
BOQ 537			1
BOQ 555		1	
BOQ 586	1		
BOQ 602	1		
BOQ 607		1	
BOQ 624		1	
BOQ 639		1	
BOQ 643			1
BOQ 672		1	
BOQ 686		1	
BOQ 693		1	
BOQ 702			1
BOQ 723		1	
Total	3	8	5

Ñandeva	AA	GG	AG
G. Ñandeva 5			1
G. Ñandeva 6		1	
G. Ñandeva 20		1	
G. Ñandeva 22		1	
G. Ñandeva 24		1	
G. Ñandeva 28		1	
G. Ñandeva 29			1
G. Ñandeva 405		1	
G. Ñandeva 825			1
G. Ñandeva 829			1
G. Ñandeva 835			1
G. Ñandeva 837	1		
G. Ñandeva 841			1
G. Ñandeva 844			1
G. Ñandeva 845		1	
G. Ñandeva 852		1	
Total	1	8	7

Kaiowa	AA	GG	AG
G. Kaiowa 203			1
G. Kaiowa 205		1	
G. Kaiowa 220		1	
G. Kaiowa 223			1
G. Kaiowa 224		1	
G. Kaiowa 229			1
G. Kaiowa 230			1
G. Kaiowa 603		1	
G. Kaiowa 608		1	
G. Kaiowa 626	1		
G. Kaiowa 634			1
G. Kaiowa 644			1

MDM2

rs2279744
chromosome 12
Chr Position 67488847

Aymaras	TT	GG	TG
BOA 158	1		
BOA 161	1		
BOA 169			1
BOA 174	1		
BOA 175			1
BOA 185	1		
BOA 204	1		
BOA 239			1
BOA 241			1
BOA 243	1		
BOA 259	1		
BOA 267	1		
BOA 269	1		
BOA 271			1
BOA 277			1
BOA 292			1
BOA 314	1		
Total	10	0	7

Quechuas	TT	GG	TG
BOQ 513			1
BOQ 516	1		
BOQ 523	1		
BOQ 529	1		
BOQ 537	1		
BOQ 555			1
BOQ 586			1
BOQ 602	1		
BOQ 607	1		
BOQ 624	1		
BOQ 639			1
BOQ 643	1		
BOQ 672	1		
BOQ 686	1		
BOQ 693			1
BOQ 702			1
BOQ 723			1
Total	10	0	7

Ñandeva	TT	GG	TG
G. Ñandeva 5		1	
G. Ñandeva 6			1
G. Ñandeva 20	1		
G. Ñandeva 22	1		
G. Ñandeva 24			1
G. Ñandeva 28	1		
G. Ñandeva 405			1
G. Ñandeva 825			1
G. Ñandeva 829			1
G. Ñandeva 835	1		
G. Ñandeva 837			1
G. Ñandeva 841			1
G. Ñandeva 844	1		
G. Ñandeva 845			1
G. Ñandeva 852	1		
Total	6	1	8

Kaiowa	TT	GG	TG
G. Kaiowa 203			1
G. Kaiowa 205	1		
G. Kaiowa 220	1		
G. Kaiowa 223	1		
G. Kaiowa 224	1		
G. Kaiowa 229	1		
G. Kaiowa 230	1		
G. Kaiowa 603			1
G. Kaiowa 608	1		
G. Kaiowa 626	1		
G. Kaiowa 634	1		
G. Kaiowa 644	1		

HAUSP/USP7

rs1529916
chromosome 16
Chr position: 8898691

Aymaras	GG	AA	GA
BOA 158	1		
BOA 161	1		
BOA 169	1		
BOA 174	1		
BOA 175			1
BOA 185	1		
BOA 204	1		
BOA 239	1		
BOA 241	1		
BOA 243			1
BOA 250	1		
BOA 259			1
BOA 267	1		
BOA 269	1		
BOA 271			1
BOA 277	1		
BOA 292	1		
BOA 314			1
Total	13	0	5

Quechuas	GG	AA	GA
BOQ 513		1	
BOQ 516	1		
BOQ 523	1		
BOQ 529	1		
BOQ 537	1		
BOQ 555			1
BOQ 586	1		
BOQ 602			1
BOQ 607	1		
BOQ 624			1
BOQ 639			1
BOQ 643	1		
BOQ 672			1
BOQ 686			1
BOQ 693	1		
BOQ 702	1		
BOQ 723	1		
Total	10	1	6

Nandeva	GG	AA	GA
G. Ñandeva 5	1		
G. Ñandeva 6			1
G. Ñandeva 20		1	
G. Ñandeva 22			1
G. Ñandeva 24			1
G. Ñandeva 28			1
G. Ñandeva 405		1	
G. Ñandeva 825		1	
G. Ñandeva 829		1	
G. Ñandeva 835			1
G. Ñandeva 837		1	
G. Ñandeva 841	1		
G. Ñandeva 844			1
G. Ñandeva 845	1		
G. Ñandeva 852			1
Total	3	5	7

Kaiowa	GG	AA	GA
G. Kaiowa 203			1
G. Kaiowa 205			1
G. Kaiowa 220		1	
G. Kaiowa 223	1		
G. Kaiowa 224			1
G. Kaiowa 229			1
G. Kaiowa 230	1		
G. Kaiowa 603	1		
G. Kaiowa 608	1		
G. Kaiowa 626	1		
G. Kaiowa 634			1
G. Kaiowa 644		1	

LIF			
G. Kaiowa 645		1	
G. Kaiowa 648	1		
G. Kaiowa 650			1
G. Kaiowa 646			1
Total	4	6	6

UROS	TT	GG	GT
U1	1		
U2	1		
U3			1
U4	1		
U5	1		
U8		1	
U10			1
U13	1		
U19	1		
U21			1
U24		1	
U26		1	
U6			1
U7			1
U9	1		
U12	1		
U14	1		
U15	1		
U22	1		
U23	1		
U25	1		
U28			1
Total	13	3	6

ANAPIA	TT	GG	GT
Ap3	1		
Ap5			1
Ap7			1
Ap18			1
Ap20			1
Ap24	1		
Ap26		1	
Ap1		1	
Ap11			1
Ap12		1	
Ap13	1		
Ap14		1	
Ap19	1		
Ap21			1
Ap23	1		
Total	5	4	6

AMANTANI	TT	GG	GT
Am1			1
Am2			1
Am3			1
Am5		1	
Am6			1
Am9	1		
Am10			1
Am11			1
Am12			1
Am13		1	
Am17		1	
Am23		1	
Am1P		1	
Am2P	1		
Am5P		1	
Am7P			1
Am13P		1	
Am7			1
Am8		1	
Am15		1	
Am18		1	
Am19		1	
Am20		1	
Am21		1	
Am8P		1	
Am9P		1	
Am10P		1	
Am11P		1	

TP53			
G. Kaiowa 645	1		
G. Kaiowa 648	1		
G. Kaiowa 650	1		
G. Kaiowa 646	1		
Total	14	0	2

UROS	CC	GG	CG
U1		1	
U2		1	
U3		1	
U4		1	
U5		1	
U8		1	
U10		1	
U13		1	
U19		1	
U21		1	
U24		1	
U26		1	
U6			1
U7		1	
U9		1	
U12			1
U14		1	
U15		1	
U22			1
U23		1	
U25		1	
U28		1	
Total		19	3

ANAPIA	CC	GG	CG
Ap3		1	
Ap5			1
Ap7		1	
Ap18			1
Ap20			1
Ap24		1	
Ap26		1	
Ap1		1	
Ap11		1	
Ap12			1
Ap13		1	
Ap14		1	
Ap19			1
Ap21			1
Ap23			1
Total		8	7

AMANTANI	CC	GG	CG
Am1			1
Am2		1	
Am3			1
Am5			1
Am6		1	
Am9		1	
Am10			1
Am11		1	
Am12		1	
Am13		1	
Am17		1	
Am23		1	
Am1P		1	
Am2P		1	
Am5P			1
Am7P			1
Am13P		1	
Am7		1	
Am8		1	
Am15		1	
Am18		1	
Am19		1	
Am20		1	
Am21		1	
Am8P		1	
Am9P		1	
Am10P		1	
Am11P		1	

MDM4			
G. Kaiowa 645			1
G. Kaiowa 648			1
G. Kaiowa 650			1
G. Kaiowa 646		1	
Total	1	6	9

UROS	AA	GG	AG
U1		1	
U2		1	
U3			1
U4		1	
U5		1	
U8		1	
U10			1
U13			1
U19			1
U21		1	
U24		1	
U26		1	
U6			1
U7		1	
U9		1	
U12		1	
U14			1
U15		1	
U22		1	
U23		1	
U25		1	
U28			1
Total		15	7

ANAPIA	AA	GG	AG
Ap3		1	
Ap5	1		
Ap7		1	
Ap18	1		
Ap20			1
Ap24			1
Ap26	1		
Ap1	1		
Ap11		1	
Ap12		1	
Ap13	1		
Ap14		1	
Ap19		1	
Ap21		1	
Ap23			1
Total	5	7	3

AMANTANI	AA	GG	AG
Am1			1
Am2			1
Am3			1
Am5		1	
Am6	1		
Am9			1
Am10			1
Am11			1
Am12		1	
Am13		1	
Am17		1	
Am23			1
Am1P			1
Am2P		1	
Am5P			1
Am7P			1
Am13P		1	
Am7			1
Am8			1
Am15	1		
Am18		1	
Am19	1		
Am20			1
Am21			1
Am8P			1
Am9P			1
Am10P			1
Am11P		1	

MDM2			
G. Kaiowa 645	1		
G. Kaiowa 648	1		
G. Kaiowa 650	1		
G. Kaiowa 646	1		
Total	14	0	2

UROS	TT	GG	TG
U1	1		
U2	1		
U3	1		
U4	1		
U5	1		
U8	1		
U10		1	
U13	1		
U19		1	
U21	1		
U24		1	
U26	1		
U6	1		
U7		1	
U9	1		
U12	1		
U14	1		
U15	1		
U22	1		
U23	1		
U25	1		
U28			
Total	17	4	1

ANAPIA	TT	GG	TG
Ap3	1		
Ap5	1		
Ap7			1
Ap18	1		
Ap20	1		
Ap24	1		
Ap26	1		
Ap1	1		
Ap11	1		
Ap12	1		
Ap13	1		
Ap14	1		
Ap19	1		
Ap21			1
Ap23	1		
Total	13	2	

AMANTANI	TT	GG	GT
Am1	1		
Am2	1		
Am3			1
Am5	1		
Am6	1		
Am9	1		
Am10	1		
Am11	1		
Am12	1		
Am13			1
Am17	1		
Am23	1		
Am1P	1		
Am2P	1		
Am5P	1		
Am7P	1		
Am13P	1		
Am7	1		
Am8			1
Am15	1		
Am18	1		
Am19	1		
Am20			1
Am21	1		
Am8P	1		
Am9P	1		
Am10P	1		
Am11P	1		

HAUSP/USP7			
G. Kaiowa 645		1	
G. Kaiowa 648	1		
G. Kaiowa 650	1		
G. Kaiowa 646	1		
Total	8	3	5

UROS	GG	AA	AG
U1	1		
U2	1		
U3	1		
U4	1		
U5	1		
U8	1		
U10	1		
U13	1		
U19	1		
U21	1		
U24			1
U26	1		
U6			1
U7	1		
U9	1		
U12	1		
U14	1		
U15	1		
U22	1		
U23	1		
U25	1		
U28	1		
Total	20		2

ANAPIA	GG	AA	AG
Ap3	1		
Ap5	1		
Ap7			1
Ap18			1
Ap20	1		
Ap24	1		
Ap26	1		
Ap1	1		
Ap11	1		
Ap12		1	
Ap13			1
Ap14		1	
Ap19	1		
Ap21		1	
Ap23	1		
Total	9	3	3

AMANTANI	GG	AA	AG
Am1	1		
Am2		1	
Am3			1
Am5	1		
Am6			1
Am9		1	
Am10			1
Am11			1
Am12			1
Am13	1		
Am17	1		
Am23	1		
Am1P	1		
Am2P	1		
Am5P	1		
Am7P			1
Am13P	1		
Am7		1	
Am8			1
Am15	1		
Am18	1		
Am19		1	
Am20			1
Am21			1
Am8P			1
Am9P		1	
Am10P			1
Am11P			1

LIF

Am12P			1
Total	2	18	9

TAQUILE	TT	GG	GT
T1		1	
T3	1		
T4		1	
T9			1
T11	1		
T13			1
T17			1
T18		1	
T19	1		
T24			1
T28			1
T29			1
T30		1	
T31			1
T32	1		
T39	1		
T40		1	
T41			1
T2P			1
T3P	1		
T4P		1	
T5P	1		
T6P			1
T7P			1
T9P	1		
T10P			1
T13P		1	
T16P			1
T17P	1		
T2			1
T14			1
T21	1		
T25		1	
T33	1		
T34	1		
T35	1		
T36		1	
T37		1	
T12P		1	
T14P	1		
T18P	1		
T19P			1
T20P	1		
Total	16	11	16

ANDOAS	TT	GG	GT
AND67			1
And-70			1
And-71			1
And-72			1
And-76			1
And-79		1	
And-83	1		
And-87			1
And-94		1	
And-95			1
And-96			1
And-98			1
And-99			1
And-104		1	
And-105		1	
And-18	1		
And-19		1	
And-24			1
And-28			1
And-34			1
And-36	1		
And-37		1	
And-38		1	
And-39			1
And-46			1
And-49			1
And-52			1

TP53

Am12P		1	
Total	1	22	6

TAQUILE	CC	GG	CG
T1		1	
T3		1	
T4		1	
T9			1
T11		1	
T13		1	
T17			1
T18		1	
T19			1
T24		1	
T28		1	
T29		1	
T30		1	
T31		1	
T32		1	
T39		1	
T40		1	
T41		1	
T2P		1	
T3P	1		
T4P		1	
T5P		1	
T6P		1	
T7P		1	
T9P		1	
T10P			1
T13P		1	
T16P		1	
T17P		1	
T2			1
T14		1	
T21		1	
T25		1	
T33		1	
T34		1	
T35		1	
T36		1	
T37		1	
T12P		1	
T14P		1	
T18P		1	
T19P		1	
T20P		1	
Total	1	37	5

ANDOAS	CC	GG	CG
And-67			1
And-70		1	
And-71	1		
And-72			1
And-76		1	
And-79			1
And-83		1	
And-87		1	
And-94		1	
And-95			1
And-96		1	
And-98		1	
And-99		1	
And-104			1
And-105		1	
And-18		1	
And-19		1	
And-24			1
And-28		1	
And-34		1	
And-36		1	
And-37			1
And-38	1		
And-39			1
And-46			1
And-49			1
And-52			1

MDM4

Am12P		1	
Total	3	9	17

TAQUILE	AA	GG	AG
T1		1	
T3			1
T4			1
T9			1
T11		1	
T13			1
T17	1		
T18		1	
T19			1
T24		1	
T28			1
T29			1
T30		1	
T31		1	
T32		1	
T39			1
T40		1	
T41		1	
T2P			1
T3P			1
T4P			1
T5P	1		
T6P			1
T7P			1
T9P		1	
T10P			1
T13P		1	
T16P			1
T17P			1
T2		1	
T14	1		
T21	1		
T25		1	
T33		1	
T34			1
T35		1	
T36			1
T37		1	
T12P			1
T14P	1		
T18P			1
T19P			1
T20P			1
Total	5	16	22

ANDOAS	AA	GG	AG
And-67		1	
And-70	1		
And-71			1
And-72			1
And-76		1	
And-79			1
And-83	1		
And-87			1
And-94		1	
And-95			1
And-96			1
And-98		1	
And-99		1	
And-104			1
And-105			1
And-18	1		
And-19			1
And-24			1
And-28	1		
And-34			1
And-36	1		
And-37	1		
And-38		1	
And-39			1
And-46			1
And-49		1	
And-52			1

MDM2

Am12P		1	
Total	25	4	

TAQUILE	TT	GG	GT
T1		1	
T3		1	
T4		1	
T9		1	
T11		1	
T13		1	
T17		1	
T18		1	
T19		1	
T24		1	
T28		1	
T29		1	
T30		1	
T31		1	
T32		1	
T39		1	
T40		1	
T41		1	
T2P		1	
T3P		1	
T4P		1	
T5P		1	
T6P		1	
T7P		1	
T9P		1	
T10P		1	
T13P		1	
T16P		1	
T17P		1	
T2		1	
T14		1	
T21		1	
T25			1
T33		1	
T34		1	
T35		1	
T36		1	
T37			1
T12P		1	
T14P		1	
T18P		1	
T19P		1	
T20P		1	
Total	40	1	2

ANDOAS	TT	GG	GT
And-67			1
And-70		1	
And-71		1	
And-72		1	
And-76			1
And-79		1	
And-83		1	
And-87		1	
And-94		1	
And-95		1	
And-96		1	
And-98		1	
And-99			1
And-104		1	
And-105		1	
And-18		1	
And-19		1	
And-24		1	
And-28		1	
And-34			1
And-36		1	
And-37		1	
And-38		1	
And-39		1	
And-46		1	
And-49		1	
And-52		1	

HAUSP/USP7

Am12P		1	
Total	11	5	13

TAQUILE	GG	AA	GA
T1	1		
T3	1		
T4	1		
T9			1
T11			1
T13			1
T17			1
T18	1		
T19	1		
T24	1		
T28	1		
T29			1
T30			1
T31			1
T32			1
T39	1		
T40	1		
T41			1
T2P	1		
T3P			1
T4P	1		
T5P	1		
T6P			1
T7P			1
T9P			1
T10P	1		
T13P			1
T16P	1		
T17P	1		
T2			1
T14			1
T21	1		
T25	1		
T33	1		
T34			1
T35	1		
T36			1
T37	1		
T12P			1
T14P	1		
T18P	1		
T19P			1
T20P	1		
Total	23		20

ANDOAS	GG	AA	GA
And-67		1	
And-70			1
And-71			1
And-72			1
And-76		1	
And-79	1		
And-83	1		
And-87			1
And-94			1
And-95		1	
And-96		1	
And-98	1		
And-99	1		
And-104	1		
And-105			1
And-18	1		
And-19		1	
And-24	1		
And-28		1	
And-34			1
And-36	1		
And-37			1
And-38	1		
And-39			1
And-46			1
And-49	1		
And-52			1

LIF			
And-58	1		
And40			1
And42		1	
And43			1
And44	1		
And45		1	
And47		1	
And48		1	
And50			1
And93			1
And97		1	
And103			1
And26		1	
And29			1
And33			1
And53		1	
And55	1		
And56			1
And57		1	
And61	1		
And63	1		
And64		1	
And66		1	
And69		1	
And73		1	
And74		1	
And77			1
And80			1
And81			1
And85			1
And88			1
And89	1		
And91		1	
And100	1		
Total	10	21	30

TP53			
And-58		1	
And40		1	
And42		1	
And43		1	
And44		1	
And45		1	
And47		1	
And48			1
And50		1	
And93	1		
And97		1	
And103		1	
And26			1
And29		1	
And33	1		
And53		1	
And55			1
And56			1
And57			1
And61		1	
And63			1
And64			1
And66		1	
And69		1	
And73		1	
And74		1	
And77		1	
And80	1		
And81			1
And85		1	
And88			1
And89		1	
And91		1	
And100	1		
Total	6	35	20

MDM4			
And-58		1	
And40	1		
And42			1
And43		1	
And44	1		
And45	1		
And47		1	
And48			1
And50	1		
And93			1
And97		1	
And103			1
And26			1
And29	1		
And33		1	
And53		1	
And55	1		
And56			1
And57			1
And61			1
And63	1		
And64	1		
And66		1	
And69			1
And73	1		
And74			1
And77			1
And80			1
And81		1	
And85			1
And88		1	
And89			1
And91		1	
And100		1	
Total	15	18	28

MDM2			
And-58	1		
And40	1		
And42	1		
And43		1	
And44		1	
And45	1		
And47	1		
And48	1		
And50	1		
And93	1		
And97	1		
And103	1		
And26	1		
And29	1		
And33	1		
And53	1		
And55		1	
And56	1		
And57		1	
And61	1		
And63	1		
And64	1		
And66	1		
And69	1		
And73	1		
And74	1		
And77	1		
And80		1	
And81		1	
And85	1		
And88	1		
And89	1		
And91		1	
And100	1		
Total	50	9	2

HAUSP/USP7			
And-58			1
And40	1		
And42	1		
And43			1
And44			1
And45			1
And47			1
And48	1		
And50			1
And93	1		
And97	1		
And103	1		
And26			1
And29		1	
And33			1
And53		1	
And55	1		
And56	1		
And57			1
And61			1
And63		1	
And64	1		
And66	1		
And69			1
And73	1		
And74			1
And77	1		
And80			1
And81			1
And85		1	
And88	1		
And89	1		
And91		1	
And100	1		
Total	25	11	25

YANQUE	TT	GG	GT
YKE1		1	
YKE2			1
YKE3			1
YKE4	1		
YKE5	1		
YKE7		1	
YKE8			1
YKE9			1
YKE10			1
YKE11	1		
Total	3	2	5

YANQUE	CC	GG	CG
YKE1		1	
YKE2			1
YKE3			1
YKE4			1
YKE5			1
YKE7		1	
YKE8			1
YKE9		1	
YKE10			1
YKE11			1
Total		3	7

YANQUE	AA	GG	AG
YKE1		1	
YKE2		1	
YKE3			1
YKE4		1	
YKE5	1		
YKE7		1	
YKE8			1
YKE9		1	
YKE10		1	
YKE11		1	
Total	1	7	2

YANQUE	TT	GG	GT
YKE1	1		
YKE2	1		
YKE3	1		
YKE4	1		
YKE5	1		
YKE7	1		
YKE8	1		
YKE9			1
YKE10	1		
YKE11	1		
Total	9	0	1

YANQUE	GG	AA	AG
YKE1	1		
YKE2	1		
YKE3	1		
YKE4			1
YKE5			1
YKE7	1		
YKE8	1		
YKE9	1		
YKE10	1		
YKE11	1		
Total	8		2

CHIVAY	TT	GG	GT
CY2			1
CY4			1
CY5			1
CY8	1		
CY9		1	
CY11			1
CY12	1		
CY13	1		
CY14	1		
CY15			1
CY16	1		
CY17		1	
CY20	1		
CY21			1
CY1			1
CY3			1
CY6			1
CY18	1		
Total	7	2	9

CHIVAY	CC	GG	CG
CY2		1	
CY4			1
CY5		1	
CY8		1	
CY9		1	
CY11		1	
CY12		1	
CY13			1
CY14			1
CY15			1
CY16		1	
CY17		1	
CY20		1	
CY21		1	
CY1		1	
CY3			1
CY6		1	
CY18			1
Total		12	6

CHIVAY	AA	GG	AG
CY2	1		
CY4			1
CY5			1
CY8			1
CY9		1	
CY11			1
CY12	1		
CY13			1
CY14	1		
CY15		1	
CY16		1	
CY17			1
CY20		1	
CY21		1	
CY1		1	
CY3		1	
CY6		1	
CY18		1	
Total	3	9	6

CHIVAY	TT	GG	GT
CY2	1		
CY4		1	
CY5	1		
CY8	1		
CY9	1		
CY11			1
CY12	1		
CY13	1		
CY14		1	
CY15	1		
CY16	1		
CY17	1		
CY20	1		
CY21	1		
CY1	1		
CY3	1		
CY6	1		
CY18	1		
Total	15	2	1

CHIVAY	GG	AA	GA
CY2	1		
CY4	1		
CY5	1		
CY8	1		
CY9	1		
CY11	1		
CY12	1		
CY13	1		
CY14	1		
CY15	1		
CY16	1		
CY17	1		
CY20	1		
CY21	1		
CY1	1		
CY3	1		
CY6	1		
CY18	1		
Total	18		

CABANACONDE	TT	GG	GT
CB1			1
CB2			1
CB3			1
CB4	1		
CB5	1		
CB6			1

CABANACONDE	CC	GG	CG
CB1		1	
CB2			1
CB3		1	
CB4		1	
CB5		1	
CB6			1

CABANACONDE	AA	GG	AG
CB1		1	
CB2			1
CB3		1	
CB4		1	
CB5			1
CB6	1		

CABANACONDE	TT	GG	GT
CB1		1	
CB2	1		
CB3		1	
CB4	1		
CB5	1		
CB6	1		

CABANACONDE	GG	AA	AG
CB1			1
CB2	1		
CB3	1		
CB4	1		
CB5	1		
CB6	1		

LIF			
CB7	1		
CB8			1
CB9			1
CB12			1
CB13			1
CB15			1
CB17	1		
CB18	1		
CB19	1		
CB20			1
CB11			1
Total	6		11

TP53			
CB7	1		
CB8	1		
CB9			1
CB12	1		
CB13	1		
CB15	1		
CB17	1		
CB18	1		
CB19	1		
CB20			1
CB11	1		
Total		13	4

MDM4			
CB7	1		
CB8			1
CB9	1		
CB12	1		
CB13			1
CB15	1		
CB17	1		
CB18	1		
CB19			1
CB20			1
CB11			1
Total	1	9	7

MDM2			
CB7	1		
CB8	1		
CB9	1		
CB12	1		
CB13	1		
CB15	1		
CB17	1		
CB18	1		
CB19	1		
CB20	1		
CB11	1		
Total	15	2	

HAUSP/USP7			
CB7	1		
CB8	1		
CB9	1		
CB12	1		
CB13	1		
CB15	1		
CB17			1
CB18	1		
CB19	1		
CB20	1		
CB11			1
Total	14		3

CAPÍTULO 4

Selection scan reveals three new loci related to high altitude adaptation in Native Andeans

Publicado como um artigo científico:**Jacovas VC,** Couto-Silva CM, Nunes K, Lemes RB, de Oliveira MZ, Salzano FM, Bortolini MC, Hünemeier T. Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. Sci Rep. 2018 Aug 24;8(1):12733. doi: 10.1038/s41598-018-31100-6. PubMed PMID: 30143708; PubMed Central PMCID: PMC6109162.

SCIENTIFIC REPORTS

OPEN

Selection scan reveals three new loci related to high altitude adaptation in Native Andeans

Received: 14 May 2018
Accepted: 6 August 2018
Published online: 24 August 2018

Vanessa C. Jacovas[1], Cainã M. Couto-Silva[2], Kelly Nunes[2], Renan B. Lemes[2], Marcelo Z. de Oliveira[3], Francisco M. Salzano[1], Maria Cátira Bortolini[1] & Tábita Hünemeier[2]

The Andean Altiplano has been occupied continuously since the late Pleistocene, ~12,000 years ago, which places the Andean natives as one of the most ancient populations living at high altitudes. In the present study, we analyzed genomic data from Native Americans living a long-time at Andean high altitude and at Amazonia and Mesoamerica lowland areas. We have identified three new candidate genes - *SP100*, *DUOX2* and *CLC* - with evidence of positive selection for altitude adaptation in Andeans. These genes are involved in the *TP53* pathway and are related to physiological routes important for high-altitude hypoxia response, such as those linked to increased angiogenesis, skeletal muscle adaptations, and immune functions at the fetus-maternal interface. Our results, combined with other studies, showed that Andeans have adapted to the Altiplano in different ways and using distinct molecular strategies as compared to those of other natives living at high altitudes.

Along their great expansion, humans have inhabited almost all environments in the five continents. Among several harsh environments that were occupied, the highlands are probably the ones that needed more adaptations for survival[1]. At least in three geographically distinct locations have this evolutionary adaptation been studied: Andean Altiplano (South America), Himalaya (China/Tibet, Asia) and Semien Mountain (northern Ethiopia, Africa) Plateaus. Andes have been peopled continuously since the late Pleistocene, ~12,000 yBP[2] while the time of settlement and permanent occupation of both Tibet and Ethiopia remain a topic of debate, varying widely[3,4]. Despite some uncertainties in the permanent occupation dating, it is certain that humans have inhabited these regions of hostile climates for thousands of years.

Several physiologic factors are associated with living at high altitude (≥2,500 meters where only 75% of the oxygen available at sea level occurs; http://www.altitude.org/air_pressure.php), including adaptations for high ultraviolet radiation index, thermal amplitude, and changes in the pulmonary capacity due to hypoxia[5,6]. High altitude leads to a rapid physiologic/adaptive response in individuals from lowlands; however, prolonged exposure to environmental-related factors might have harmful outcomes. Remarkable features such as increased pulmonary function, hypoxia tolerance, and increased hemoglobin levels have been observed in Andean populations[7]. How such adaptations took place is still not clear, and just a few genes have been associated with the high altitude adaptation phenotype in human populations[8–13].

Interestingly, the set of genes presenting signs of natural selection changes according to high altitude, indicating that under an analogous selective pressure, different genetic solutions have emerged. For instance, genomic scans for selection have revealed at least 40 candidate genes related to the Hypoxia Inducible Factor (HIF), such as *EPAS1* in populations from Tibet, *EGLN1* in Andeans and Tibetans and *THRB* and *ARNT2* in Ethiopians[8,14–18]. The populations from the Andean plateau also presented signs of natural selection in other genes, such as *BRINP3*, *NOS2*, and *TBX5*, involved in the nitric oxide pathway (NOS) and related to cardiovascular health[12]. In addition, Jacovas *et al.*[19] using the candidate gene approach inferred that a combination of some derived and ancestral alleles of *USP7, LIF* and *MDM2* genes, all three in the *TP53* pathway, could have been essential for the successful establishment of Native American populations in the Andean highlands.

[1]Genetics Departament, Biosciences Institute, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil. [2]Department of Genetics and Evolutionary Biology, Biosciences Institute, University of São Paulo, São Paulo, SP, Brazil. [3]Vale do Rio dos Sinos University, São Leopoldo, RS, Brazil. Vanessa C. Jacovas, Cainã M. Couto-Silva and Kelly Nunes contributed equally. Correspondence and requests for materials should be addressed to M.C.B. (email: maria.bortolini@ufrgs.br) or T.H. (email: hunemeier@usp.br)

SNP	Alelle		Gene	Position	PBS	XPEHH			
	Ancestral	Derived				Andean *vs.* Mesoamerican	*p*-value	Andean *vs.* Amazonian	*p*-value
rs13411586	C*	T	*SP100*	230988046	0.5846	2.3789	0.0037	2.1703	0.0065
rs9678342	C*	T	*SP100*	230991955	0.5547	2.3193	0.0044	2.1356	0.0071
rs7582700	T*	C	*SP100*	231024349	0.4644	2.2704	0.0050	2.1074	0.0076
rs7039618	A	G*	*TMEM38B*	107497627	0.3618	0.0842	0.3312	0.5672	0.1458
rs3817141	T*	C	*TMEM38B*	107507950	0.3906	0.0255	0.3099	0.6205	0.1351
rs10978213	G*	A	*TMEM38B*	107511706	0.3618	0.0235	0.3092	0.6171	0.1358
rs10816302	A*	G	*TMEM38B*	107526354	0.3835	0.0937	0.2697	0.6664	0.1264
rs10978240	A	G*	*TMEM38B*	107575093	0.3923	0.0764	0.2753	0.6307	0.1331
rs1046778	T	C*	*AS3MT*	104651474	0.3124	0.5023	0.5118	0.4008	0.4631
rs269866	G*	A	*DUOX2*	43181698	0.6185	2.0599	0.0086	2.5865	0.0021
rs440191	A	G*	*CLC*	44913483	0.3039	1.6207	0.0234	0.3166	0.2046

Table 1. Population Branch Statistic (PBS) individual values and Cross-Population Extended Haplotype Homozygosity (XP-EHH) for all SNPs found under selection in Native Andean populations. Ancestral and derived alleles according to the 1000 Genomes data. *Putative selected alleles.

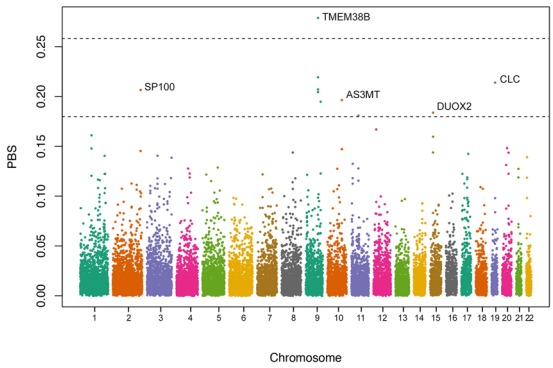


Figure 1. Average PBS values in windows of 20 SNPs, using a step size of 5 SNPs. The 99.5th and 99.9th percentiles of the empirical distribution are shown as black dashed horizontal lines. Names of genes associated with the highest peaks are shown.

Since different investigations pointed to distinct sets of genes involved in high altitude adaptation, more studies are necessary to fully understand the different genetic landscapes present in highland populations around de world. In the present study, we compared genomic data from Native American populations living for a long-time at high altitude (Andean Altiplano) with those living at lowlands (Amazon and Mesoamerica), with the purpose of expanding our knowledge about the genetic repertoire responsible for the successful human colonization of the Andes.

Results

Natural selection analysis. Population Branch Statistic (PBS) values were estimated for each individual SNP. To avoid spurious results due to single SNPs, windows of 20 SNPs were used to estimate the mean PBS values for a given region. Then, we checked the outliers' peaks, above the 99.5th and 99.9th percentiles, to identify in each outlier window the SNPs with the highest PBS value and assigned the gene to which it belonged (or the nearest gene). Based on this approach, five candidate genes were identified: *SP100* (SP100 Nuclear Antigen), *TMEM38B* (Transmembrane Protein 38B), *AS3MT* (Arsenite 3 Methyltransferase), *DUOX2* (Dual Oxidase 2) and *CLC* (Charcot-Leyden Crystal Galectin, also known as Galectin-10) (Table 1 and Fig. 1). Among these candidate genes, *AS3MT and TMEM38B* have been identified in previous scans for natural selection in Andeans[13,20].

Neutral coalescent simulations indicated that these deviations were statistically significant (p-values ranging between 0.03 and 0.0001; Fig. 2, Table S1), consistent with the action of positive selection as opposed to genetic drift in increasing the frequency of the putative selected alleles at all five tested loci. In addition, we applied the Cross-Population Extended Haplotype Homozygosity (XP-EHH) test to the same regions. The XP-EHH results also show significant differences between the Andean and Mesoamerican groups in three SNPs (rs13411586, rs9678342, rs7582700) of *SP100* and one SNP (rs269866) of *DUOX2* (Table 1). These SNPs, which are under putative selection in the PBS analysis with the most extreme values (0.46 to 0.62), also present significant XP-EHH values ≥ 2 in both Andean vs Mesoamerican and Andean vs Amazonian groups.

The observed allele density provided by the iHS test showed a notable Gaussian distribution pattern for all three groups (Fig. S1), with homozygosity decaying according to the distance from the focal markers.

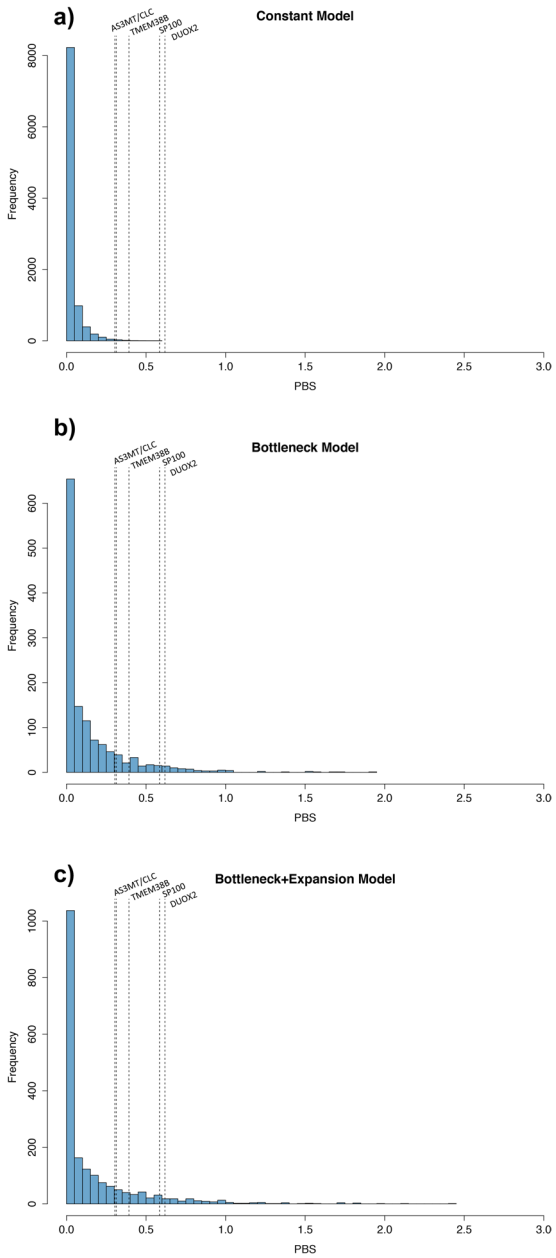


Figure 2. Distribution of 10,000 simulated PBS values under three neutral coalescent models. (**a**) Constant population model. (**b**) Population bottleneck model; and (**c**) Population bottleneck followed by expansion model. The dashed line represents the top observed PBS SNP values in the empirical datasets.

It should be noted that the distribution of alleles C (rs13411586, *SP100*), G (rs269866, *DUOX2*) and G (rs440191, *CLC*), which presented the highest PBS values (Table 1), showed their highest values in areas of very high Andean altitudes (Table 2 and Fig. 3).

Bootstrap simulations indicated that in all instances the 95% confidence interval of allele frequencies in lowlanders does not include the average values observed for populations living in high altitudes (>4000 m above sea level) (Fig. S2), suggesting that the differences found in allele frequencies between population groups might be caused by a non-random evolutionary process.

Effects of putatively selected alleles on gene expression. Homozygotes for the *DUOX2* putatively selected allele (rs269866 G) presented a slight increase in the expression of the DUOX2 protein (Fig. S3). Multiple testing across tissues showed significant expression of this protein in thyroid (m-value = 1.0), lungs (m-value = 0.996) and aorta artery (0.996) (Fig. S4). Homozygotes for the rs13411586 (*SP100*) putatively selected allele (C) presented an increase in the expression of the SP100 protein in skeletal muscles (Fig. S5). Multiple testing across tissues showed significant expression of this protein in skeletal muscle (m-value = 1.0) and testis (m-value = 0.971) (Fig. S6). There is no information available about the *CLC* gene expression profile.

Population (n)	*DUOX2* G allele (rs269866)	*SP100* C allele (rs13411586)	*CLC* G allele (rs440191)
Mesoamerican Lowland (<2,500 m.)			
Total (153)	0.068*	0.045*	0.128*
South American (Andean) Highland (≥4,000 m.)			
Total (63)	0.420*	0.397*	0.452*
South American (Amazonian) Lowland (<2,500 m.)			
Total (106)	0.048*	0.053*	0.142*

Table 2. Frequencies of the putatively selected alleles in the populational groups. *Weighted average.

Discussion

We identified five loci under positive selection in Andean Native populations. Two of them were previously described: *AS3MT* was found to be under positive selection in Colla Andeans systematically exposed to arsenic water[20] while *TMEM38B* reduced the negative effects of polycythemia (elevated hematocrit or decreased plasma volume) at high altitudes[13]. Three other genes, *SP100*, *DUOX2*, and *CLC* were identified for the first time in a high-altitude context in the present study. These genes are part of the *TP53* pathway, already indicated as a potential candidate to be under natural selection in high altitude populations[19,21].

SP100 is a single-copy gene in the human genome that produces several alternatively spliced Sp100 protein isoforms known as modulators of the p53 activity[22]. We found three SNPs in the *SP100* gene with high and significant PBS values, as well as significant XP-EHH values when Andeans were compared to others. One of these SNPs, rs13411586, is differentially expressed in humans; our prediction showed that individuals homozygous for the putatively selected allele (C) have increased Sp100 production.

Interestingly, we also identified that the *SP100* gene is differentially expressed in skeletal muscles (Fig. S3). Studies have revealed that a member of the HIF pathway, HIF-1, plays an important role in the regulation of oxygen homeostasis, which includes the physiological skeletal and heart muscle adaptations in situations of oxygen reduction due to muscular effort[23–25] and ischemic cardiomyopathy, respectively[26]. Exposure to high altitude leads to reduced muscle mass and performance (*e.g.* lower work capacity and standing fatigue), except when one is evolutionarily adapted to it[27–29].

HIF-1 protects cell-survival during low oxygen supply, while p53 promotes genome cell-death under hypoxia. The reason for these apparently antagonistic roles can be in the difference of the oxygen quantity available; in a normal condition, both p53 and HIF-1 levels are low, but in mild hypoxia, the p53 level remains low, whereas the HIF-1 level increases, protecting cells still relatively healthy from destruction. In severe hypoxia, p53 accumulation promotes the repression or degradation of anti-apoptotic proteins like HIF-1, inducing apoptosis of the damaged cells[30–32]. Sp100 is known as a modulator of the p53 activity[22] and under tissue hypoxia due to ischemia, it is downregulated, leading to genomic instability[26]. The Andean population presents high allele C (rs13411586) frequency (Table 2), which in homozygosis increase Sp100 production according to our prediction test. Our result suggests an evolutionary solution to keep Sp100 at an adequate level in an environment with a constant low oxygen level. Furthermore, it is possible to speculate that there is an intricate balance in the level of expression of the *SP100, TP53* and *HIF-1* genes under hypoxia, considering both short (reversible physiological and metabolic adaptations) and long-term evolutionary adaptation scenarios.

DUOX2, expressed in epithelial cells of various tissues including nasal and lung, participates in the hydrogen peroxide (H_2O_2) pathway, which is required in the final steps of thyroid hormones production. It is also involved in Reactive Oxygen Species (ROS), a byproduct of the normal oxygen metabolism even under normal physiologic conditions[33]. However, different stressor conditions can increase the ROS production, *i.e.* high-altitude exposure (hypoxia and UV exposure), and pathological conditions such as cancer[34]. Salmeen *et al.*[35] provided evidence that DUOX2 plays a role in a p53-dependent checkpoint mechanism for cell cycle entry.

In vitro and *in vivo* experiments showed that oxidative stress and generation of ROS caused by DUOX2 overexpression, in both hypoxia and hyperoxia, contribute to inflammation, carcinogenesis and cell death[36–42]. For instance, a functional study[36] showed that under hyperoxia conditions, mutant mice for DUOX2 had significant lower acute lung injuries induced by hyperoxia. This finding pointed to the importance of these proteins in the response to changes of oxygen concentration in the environment. Another study[37] found that chickens submitted to hypoxia (>3,000 m) had increased activity of DUOX/NOX proteins, indicating the physiological role of these enzymes in the process of adaptation to oxidative stress.

Our results on the expression of the *DUOX2* putatively selected allele G (high PBS values and significant XP-EHH value > 2; Table 1) also pointed to higher levels of protein expression in humans, mainly in the lungs and arteries. It is noteworthy that ROS contributes to inflammation in the vessel walls. Kim & Byzowa[43] demonstrated that ROS has an important role in angiogenesis, a process of new blood vessel growth. Angiogenesis is a key event in the physiological response to hypoxia and therefore might have a role in the adaptation to high altitude in long-term residents, especially in individuals with excessive erythropoiesis (like those found in the Chronic Mountain Sickness [CMS] phenotype), to compensate a plausible change in microcirculation[44,45].

SNP rs440191 is located at the 3′UTR region of *CLC*, and the putatively selected allele G is in complete linkage disequilibrium with the *CLC* rs395892 G allele in the Mexican population[46]. The latter is associated with eosinophil and basophil counts[47], while rs440191 has so far been investigated just in approaches assessing allergic susceptibilities[48]. Gene expression queries did not show any significant eQTL related to this polymorphism, preventing any prediction of tissue-specific expression.

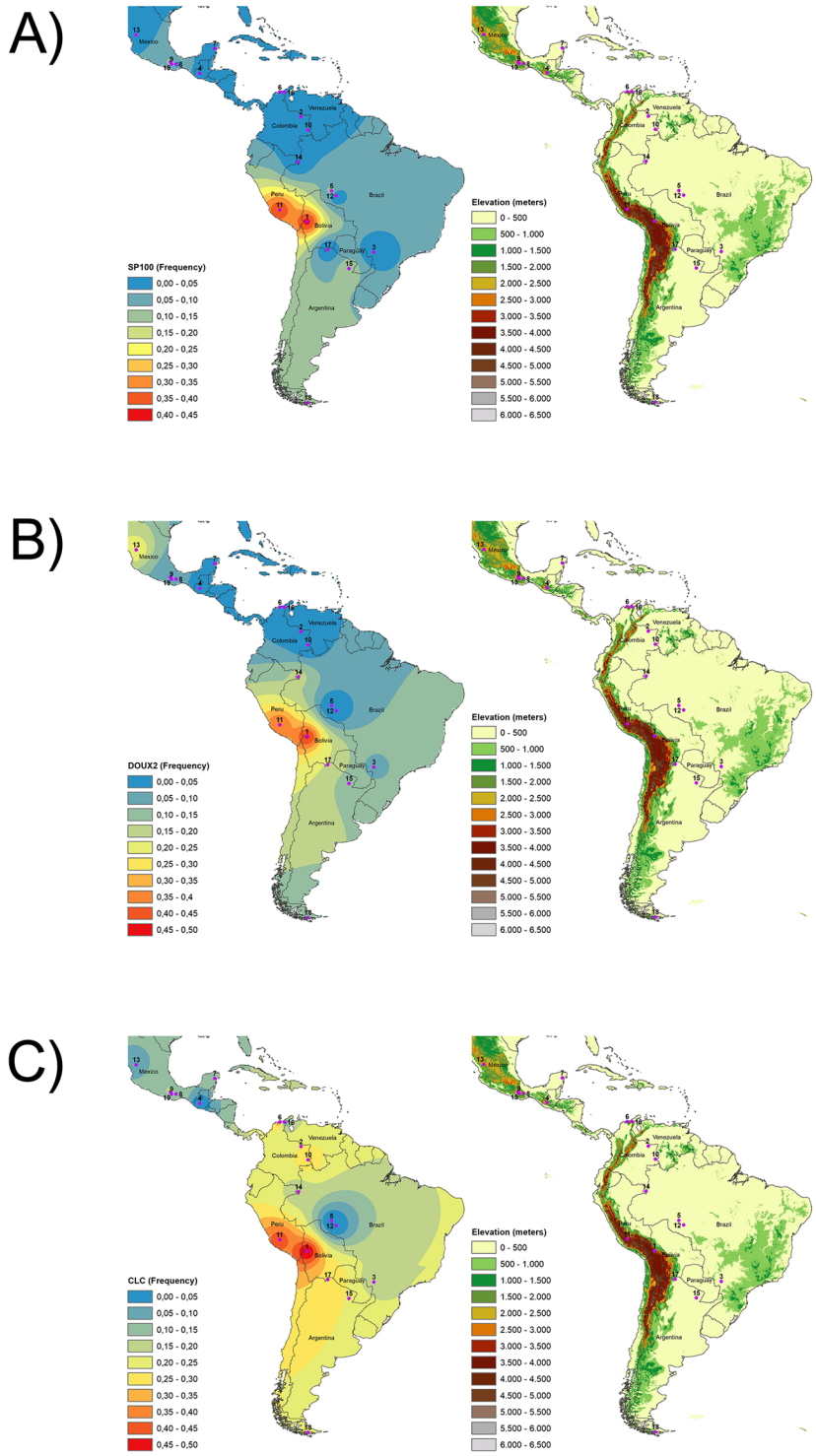


Figure 3. (**a**) rs13411586_C (*SP100*). (**b**) rs269866_G (*DUOX2*) and (**c**) rs440191_A (*CLC*) allele frequency distributions according to altitude. Populations ($n \geq 3$): 1. Aymara, 2. Guahibo, 3. Guarani, 4. Kaqchikel, 5. Karitiana, 6. Kogi, 7. Maya, 8. Mixe, 9. Mixtec, 10. Piapoco, 11. Quechua, 12. Surui,13. Tepehuano, 14. Ticuna, 15. Toba, 16. Wayuu, 17. Wichi, 18. Yaghan and 19. Zapotec.

CLC (galectin-10) is still a poorly studied gene when compared to other members of the functionally polyvalent galectin family. It is recognized as a lysophospholipase expressed in eosinophils and basophils, although some authors identified it just as an enzyme that interacts with lysophospholipases[49]. The only functional study regarding this protein showed that hypoxia increases eosinophil accumulation and CLC production in humans, concomitant with a delay in constitutive apoptosis, antagonizing the normal pro-apoptotic effect of agents that normally induce eosinophil apoptosis[50].

Regulation by the p53 transcription factor seems to be important in the galectin family genes' expression. For instance, the galectin-3 gene has a binding site for p53, and p53 increases the transcription of paralogue galectin-7[51–53]. Altered expression of galectin genes, including *CLC*, was implicated in cancer emergence and progression, highlighting the role of the galectins in cell proliferation via cell death programs[54].

Investigations with galectin paralogues have shown that galectin-1 in the first term ovine gestation placenta prevented inflammatory processes that harm the fetus[55], while galectin-13, which has the highest homology to CLC, is a member of the group of the so-called "pregnancy-related proteins", due to its special immune functions at the feto-maternal interface[56,57]. These fundamental cell functions, already described for humans and other placental mammals, may indicate the path that connects our CLC findings and the selection pressure in the Andean hostile climate.

In conclusion, our results pointed to a complex adaptation that occurred in Andean natives, which involved the *CLC, SP100* and *DUOX2* genes, not previously correlated in contexts of long-time adaptation to high altitudes. We also reinforced the role of the *TP53* pathway at least for the adaptation to the Andean environmental stresses. Combined with other studies, and incorporating the present one, it is clear that Andeans have adapted to the Altiplano in different ways and using distinct molecular strategies than those of other natives living at high altitude.

Methods

Populations. We analyzed 213,987 SNPs determined with Illumina 610quad from 63 Native Americans living at extreme high altitude ($\geq$4,000 m; 63% of the oxygen available at sea level; http://www.altitude.org/air_pressure.php) and 259 living at lowland areas ($<$2,500 m), data previously published by Reich *et al.*[58]. Highlanders included Aymara and Quechua Andeans, while lowlanders were represented by 25 populations from the Mesoamerican and South American lowlands. Details about these populations, sample sizes and allelic frequencies are given in Table S2. Additional information, including ethical authorizations for evolutionary and anthropological studies, can be found in the primary publication[58].

Population Branch Statistic (PBS) analysis. PBS determinations were performed between pair of populations, using Andean and Amazonian populations as sister groups and Mesoamericans as an outgroup. The analysis was carried out as described by Yi *et al.*[59], with only the polymorphic SNPs in at least two of the populations being considered. From the genetic distances (F_{ST}) between the three population groups examined, PBS measures if there are alleles with extreme frequencies in the Andean group as compared to the other two. Under a scenario of genetic drift only, we expect that Andeans and Amazonians will be more similar genetically than both compared to Mesoamericans. If, however, there has been local adaptation, we should detect genes that have been targeted by selection in Andeans. PBS values were estimated for both individual SNPs and windows of 20 SNPs overlapped in five SNPs. The empiric distribution of PBS values, with a 99.5th threshold, was used to determine signals of positive selection (more details in Amorim *et al.*[60]).

Demographic simulations. To verify the significance of the observed positive selection signals we simulated different demographic models, according to reported historical population data and inferred effective population sizes. We adapted the models described by Valverde *et al.*[11], to account for the divergence between Mesoamericans, Andeans and Amazonians. Assuming that the American continent was peopled beginning at 15,000 yBP, the Andes colonized by 12,000 yBP and the Amazon by 10,000 yBP, and based on Nes estimated by Valverde *et al.*[11], we simulated the three demographic models proposed by them: (a) Constant Model: Ne of 7,000 individuals with constant size in all populations throughout history; (b) Bottleneck Model: Ne 8,000 in Mesoamerica, 4,000 in Andes and 2,000 in Amazon; and (c) Bottleneck + Expansion Model: model b with bottlenecks reducing the effective size of all populations by 50% in the last 10,000 years followed by a sharp expansion in the last 8,000 years. Simulations were performed in the MS program[61] with 10,000 replicates for each demographic scenario.

Linkage disequilibrium analysis. We also used three linkage disequilibrium-based methods: extended haplotype homozygosity (EHH)[62], integrated haplotype score (iHS)[63], and cross-population extended haplotype homozygosity (XP-EHH)[64]. These approaches adopt the same core principle, that an advantageous allele under a hard sweep rise in frequency — carrying its neighbor alleles and therefore promoting homozygosity extension — quickly enough that recombination is not able to break down the haplotype. EHH statistics calculate the homozygosity rate from a core region (putative allele under selection) to the neutral scenery, *i.e.* the probability that any two randomly chosen chromosomes will be identical by descent, from the core region to a distance x. iHS evaluates the EHH considering both ancestral and derived alleles, and XP-EHH is used to calculate EHH/iHS between populations, therefore controlling for local variation. These tests are complementary; while iHS is better for detecting incomplete sweeps, XP-EHH has more power to detect sweeps near fixation[65]. Both measurements and significance were calculated through the 'rehh' R package[66].

Geographical analysis. To evaluate the variants spatial distribution, weighted inverse distance interpolation (IDW) was used to determine cell values using a weighted linear combination of a set of sample points. Weight is a function of the inverse distance[67]. The maps were made with the ArcGis 10.5 software and the cartographic base was georeferenced to the World Geodetic System (WGS84).

Bootstrap Simulations. To verify whether the allele frequencies of the candidate variants under selection are significantly different among extreme high ($>$4,000 m) and lowland ($<$4,000 m) populations, we obtained the 95% confidence intervals of the average allele frequency of the lowland populations by means of 10,000 computer-assisted bootstrap simulations with replacement, considering a sample as having the same size and

genotypic proportions observed in the real one. The average allele frequencies from high and lowland populations were obtained by weighing the observed frequencies according to their sample sizes.

Analysis of gene expression. We used the Genotype-Tissue Expression Portal (GTEx; https://www.gtex-portal.org/home/) to evaluate possible associations between each of the candidate alleles with highest differentiation and gene expression across human tissues looking for evidence of quantitative trait loci (eQTLs). The m-value is the posterior probability that an eQTL effect exists in each tissue tested in the cross-tissue meta-analysis. The m-value ranges between 0 and 1 (m-values > 0.9 mean that the tissue is predicted to have an eQTL effect).

References

1. Espinoza-Navarro, O., Díaz, J., Rodríguez, H. & Moreno, A. Effects of altitude on anthropometric and physiological patterns in Aymara and non-Aymara population between 18 and 65 years in the province of Parinacota Chile (3.700 masl). *Int. J. Morphol* **29**, 34–40 (2011).
2. Rademaker, K. *et al*. Paleoindian settlement of the high-altitude Peruvian Andes. *Science* **346**, 466–469 (2014).
3. Aldenderfer, M. Peopling the Tibetan plateau: Insights from archaeology. *High Alt. Med. Biol.* **12**, 141–147 (2011).
4. Lu, D. *et al*. Ancestral origins and genetic history of Tibetan highlanders. *Am. J. Hum. Genet.* **99**, 580–594 (2016).
5. Moore, L. G. Human genetic adaptation to high altitude. *High Alt. Med. Biol.* **2**, 257–279 (2001).
6. Beall, C. M. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl. Acad. Sci.* **104**, 8655–8660 (2007).
7. Bigham, A. W. & Lee, F. S. Human high-altitude adaptation: Forward genetics meets the HIF pathway. *Genes Dev.* **28**, 2189–2204 (2014).
8. Scheinfeldt, L. B. *et al*. Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biol.* **13**, R1, https://doi.org/10.1186/gb-2012-13-1-r1 (2012).
9. Huerta-Sánchez, E. *et al*. Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. *Mol. Biol. Evol.* **30**, 1877–1888 (2013).
10. Simonson, T. S., Huff, C. D., Witherspoon, D. J., Prchal, J. T. & Jorde, L. B. Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans. *Exp. Physiol.* **100**, 1263–1268 (2015).
11. Valverde, G. *et al*. A novel candidate region for genetic adaptation to high altitude in Andean populations. *PLoS One* **10**, e0125444 (2015).
12. Fehren-Schmitz, L. & Georges, L. Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian highlanders in the last 8500 years. *Sci. Rep.* **6**, 23485 (2016).
13. Crawford, J. E. *et al*. Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *Am. J. Hum. Genet.* **101**, 752–767 (2017).
14. Beall, C. M. *et al*. Natural selection on EPAS1 (HIF2) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl. Acad. Sci.* **107**, 11459–11464 (2010).
15. Bigham, A. *et al*. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet*. **6** (2010).
16. Simonson, T. S. *et al*. Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**, 72–75 (2010).
17. Peng, Y. *et al*. Genetic variations in tibetan populations and high-altitude adaptation at the Himalayas. *Mol. Biol. Evol.* **28**, 1075–1081 (2011).
18. Xu, S. *et al*. A genome-wide search for signals of high-altitude adaptation in tibetans. *Mol. Biol. Evol.* **28**, 1003–1011 (2011).
19. Jacovas, V. C. *et al*. Genetic variations in the TP53 pathway in native americans strongly suggest adaptation to the high altitudes of the Andes. *PLoS One* **10**, e0137823 (2015).
20. Eichstaedt, C. A. *et al*. Positive selection of AS3MT to arsenic water in Andean populations. *Mutat. Res. Mol. Mech. Mutagen.* **780**, 97–102 (2015).
21. Eichstaedt, C. A. *et al*. The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. *PLoS One* **9**, e93314 (2014).
22. Berscheminski, J. *et al*. Sp100A is a tumor suppressor that activates p53-dependent transcription and counteracts E1A/E1B-55K-mediated transformation. *Oncogene* **35**, 3178–3189 (2016).
23. Semenza, G. L. Regulation of mammalian O2 homeostasis by hypoxia-inducible Factor 1. *Annu. Rev. Cell Dev. Biol.* **15**, 551–578 (1999).
24. Vogt, M., Billeter, R. & Hoppeler, H. Effect of hypoxia on muscular performance capacity: Living low-training high. *Ther. Umsch.* **60**, 419–424 (2003).
25. Lindholm, M. E. & Rundqvist, H. Skeletal muscle hypoxia-inducible factor-1 and exercise. *Exp. Physiol.* **101**, 28–32 (2016).
26. Herrer, I. *et al*. Gene expression network analysis reveals new transcriptional regulators as novel factors in human ischemic cardiomyopathy. *BMC Med. Genomics* **8**, 14 (2015).
27. Marconi, C., Marzorati, M. & Cerretelli, P. Work capacity of permanent residents of high altitude. *High Alt. Med. Biol.* **7**, 105–115 (2006).
28. Brutsaert, T. D. Do high-altitude natives have enhanced exercise performance at altitude? *Appl. Physiol. Nutr. Metab.* **33**, 582–592 (2008).
29. Counter, S. A., Buchanan, L. H., Ortega, F., Jacobs, A. B. & Laurell, G. Assessment of the brainstem-mediated stapedius muscle reflex in Andean children living at high altitudes. *High Alt. Med. Biol.* **18**, 37–45 (2017).
30. Schmid, T., Zhou, J. & Brüne, B. HIF-1 andp53: communication of transcription factors under hypoxia. *J. Cell. Mol. Med.* **8**, 423–431 (2004).
31. Obacz, J., Pastorekova, S., Vojtesek, B. & Hrstka, R. Cross-talk between HIF and p53 as mediators of molecular responses to physiological and genotoxic stresses. *Mol. Cancer* **12**, 93 (2013).
32. Zhou, C.-H., Zhang, X.-P., Liu, F. & Wang, W. Modeling the interplay between the HIF-1 and p53 pathways in hypoxia. *Sci. Rep.* **5**, 13834 (2015).
33. Devasagayam, T. P. A. *et al*. Free radicals and antioxidants in human health: current status and future prospects. *J. Assoc. Physicians India* **52**, 794–804 (2004).
34. Gupta, S. C. *et al*. Upsides and downsides of reactive oxygen species for cancer: the roles of reactive oxygen species in tumorigenesis, prevention, and therapy. *Antioxid. Redox Signal.* **16**, 1295–1322 (2012).
35. Salmeen, A., Park, B. O. & Meyer, T. The NADPH oxidases NOX4 and DUOX2 regulate cell cycle entry via a p53-dependent pathway. *Oncogene* **29**, 4473–4484 (2010).
36. Kim, M.-J. *et al*. Dual oxidase 2 in lung epithelia is essential for hyperoxia-induced acute lung injury in mice. *Antioxid. Redox Signal.* **21**, 1803–18 (2014).
37. Bautista-Ortega, J., Cortes-Cuevas, A., Ellis, E. A. & Ruiz-Feria, C. A. Supplemental L-arginine and vitamins e and C preserve xanthine oxidase activity in the lung of broiler chickens grown under hypobaric hypoxia. *Poult. Sci.* **93**, 979–988 (2014).

38. Fletcher, N. M. *et al*. Nicotinamide adenine dinucleotide phosphate oxidase is differentially regulated in normal myometrium versus leiomyoma. *Reprod. Sci.* **21**, 1145–1152 (2014).
39. Dias-Freitas, F., Metelo-Coimbra, C. & Roncon-Albuquerque, R. Molecular mechanisms underlying hyperoxia acute lung injury. *Respir. Med.* **119**, 23–28 (2016).
40. Min, H. J. *et al*. ROS-dependent HMGB1 secretion upregulates IL-8 in upper airway epithelial cells under hypoxic condition. *Mucosal Immunol.* **10**, 685–694 (2017).
41. Lin, S.-C. *et al*. High immunoreactivity of DUOX2 is associated with poor response to preoperative chemoradiation therapy and worse prognosis in rectal cancers. *J. Cancer* **8**, 2756–2764 (2017).
42. MacFie, T. S. *et al*. DUOX2 and DUOXA2 form the predominant enzyme system capable of producing the reactive oxygen species H2O2in active ulcerative colitis and are modulated by 5-aminosalicylic acid. *Inflamm. Bowel Dis.* **20**, 514–524 (2014).
43. Kim, Y. & Byzova, T. V. Oxidative stress in angiogenesis and vascular disease. *Blood* **123**, 625–632 (2015).
44. Ge, R.-L. *et al*. B-type natriuretic peptide, vascular endothelial growth factor, endothelin-1, and nitric oxide synthase in chronic mountain sickness. *Am. J. Physiol. Heart Circ. Physiol.* **300**, H1427–33 (2011).
45. Buroker, N. E. *et al*. AKT3, ANGPTL4, eNOS3, and VEGFA associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. *Int. J. Hematol.* **96**, 200–213 (2012).
46. The 1000 Genomes Project. http://www.nature.com/articles.nature09534.
47. Astle, W. J. *et al*. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 (2016).
48. Bryborn, M., Halldén, C., Säll, T. & Cardell, L. O. CLC - a novel susceptibility gene for allergic rhinitis? *Allergy* **65**, 220–228 (2010).
49. Ackerman, S. J. *et al*. Charcot-leyden crystal protein (galectin-10) is not a dual function galectin with lysophospholipase activity but binds a lysophospholipase inhibitor in a novel structural fashion. *J. Biol. Chem.* **277**, 14859–14868 (2002).
50. Porter, L. M. *et al*. Hypoxia causes IL-8 secretion, Charcot Leyden crystal formation, and suppression of corticosteroid-induced apoptosis in human eosinophils. *Clin. Exp. Allergy* **47**, 770–784 (2017).
51. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. A model for p53-induced apoptosis. *Nature* **389**, 300–305 (1997).
52. Raimond, J., Rouleux, F., Monsigny, M. & Legrand, A. The second intron of the human galectin-3 gene has a strong promoter activity down-regulated by p53. *FEBS Lett.* **363**, 165–169 (1995).
53. Cooper, D. N. W. Galectinomics: finding themes in complexity. *Biochim. Biophys. Acta - Gen. Subj.* **1572**, 209–231 (2002).
54. Gopalan, V. *et al*. The expression profiles of the galectin gene family in colorectal adenocarcinomas. *Hum. Pathol.* **53**, 105–113 (2016).
55. Iglesias, M. M., Rabinovich, G. A., Ivanovic, V., Sotomayor, C. & Wolfenstein-Todel, C. Galectin-1 from ovine placenta. *Amino-acid sequence, physicochemical properties and implications in T-cell death. Eur. J. Biochem.* **252**, 400–407 (1998).
56. Than, N. G. *et al*. Functional analyses of placental protein 13/galectin-13. *Eur. J. Biochem.* **271**, 1065–1078 (2004).
57. Su, J. *et al*. Galectin-13, a different prototype galectin, does not bind β-galacto-sides and forms dimers via intermolecular disulfide bridges between Cys-136 and Cys-138. *Sci. Rep.* **8**, 980 (2018).
58. Reich, D. *et al*. Reconstructing native American population history. *Nature* **488**, 370–374 (2012).
59. Yi, X. *et al*. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
60. Amorim, C. E. *et al*. Genetic signature of natural selection in first Americans. *Proc. Natl. Acad. Sci.* **114**, 2195–2199 (2017).
61. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).
62. Sabeti, P. C. *et al*. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
63. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
64. Sabeti, P. C. *et al*. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
65. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic data. *Annu. Rev. Genet* **47**, 97–120 (2013).
66. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
67. Watson, D. F. & Philip, G. M. A refinement of inverse distance weighted interpolation. *Geoprocessing* **2**, 315–327 (1985).

Acknowledgements

We thank Eduardo Gameiro for the technical support. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (Grant 15/26875-9; TH), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES(Grant1645581; KN) and Conselho Nacional de Desenvolvimento Científico e Tecnológico -CNPq.

Author Contributions

M.C.B. and T.H. conceived the study. T.H. designed the analyses. C.M.C.-S. and K.N. performed the demographic and selection analyses. V.C.J. compiled the environmental variables and populational data. R.B.L. performed the bootstrap simulations. M.Z.O. carried out the geographical analysis. F.M.S., M.C.B. and T.H. wrote the manuscript with inputs from the other authors.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-31100-6.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Material

Selection scan reveals three new loci related to high altitude adaptation in Native Andeans

Vanessa C. Jacovas[1†], Cainã M. Couto-Silva[2†], Kelly Nunes[2†], Renan B. Lemes[2], Marcelo Z. de Oliveira[3], Francisco M. Salzano[1], Maria Cátira Bortolini[1*] & Tábita Hünemeier[2*]

[1]Genetics Departament, Biosciences Institute, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil.

[2]Department of Genetics and Evolutionary Biology, Biosciences Institute, University of São Paulo, São Paulo, SP, Brazil.

[3]Vale do Rio dos Sinos University, São Leopoldo, RS, Brazil.

[†]These authors contributed equally to this work.

*Correspondence and requests for materials should be addressed to M.C.B. (email: maria.bortolini@ufrgs.br) or T.H. (email: hunemeier@usp.br)

Table S1. Significance of the PBS values for the extreme SNPs of each candidate gene, obtained for each simulated demographic model.

SNP (Gene)	Constant Model	Bottleneck+Expansion	Bottleneck
rs13411586 (*SP100*)	0.0001	0.0143	0.0074
rs10978240 (*TMEM38B*)	0.0017	0.0248	0.0147
rs1046778 (*AS3MT*)	0.0047	0.0315	0.0192
rs269866 (*DUOX2*)	0.0001	0.0132	0.0060
rs440191 (*CLC*)	0.0056	0.0321	0.0206

Table S2. Allelic frequencies by Native American population analyzed in the present study (*Weighted average).

Population (n)	*DUOX2* G allele (rs269866)	*SP100* C allele (rs13411586)	*CLC* G allele (rs440191)
Mesoamerican Lowland (< 2,500 m.)			
Kaqchikel (13)	0	0.042	0
Maya (49)	0	0	0.138
Mixe (17)	0.059	0.029	0.147
Mixtec (5)	0.100	0.100	0.100
Purepecha (1)	0	0	0
Tepehuano (25)	0.240	0.020	0.080
Zapotec (43)	0.068	0.114	0.182
Total (153)	**0.068***	**0.045***	**0.128***
South American (Andean) Highland (≥ 4,000 m.)			
Aymara (23)	0.457	0.413	0.500
Quechua (40)	0.400	0.388	0.425
Total (63)	**0.420***	**0.397***	**0.452***
South American (Amazonian) Lowland (< 2,500 m.)			
Guahibo (6)	0	0	0.250
Guarani (6)	0.083	0	0.167
Jamamadi (1)	0	0.500	1
Kaingang (2)	0.500	0	0
Karitiana (13)	0	0.115	0
Kogi (4)	0	0	0.375
Maleku (3)	0	0	0
Palikur (3)	0	0.167	0.500
Parakana (1)	0	0.500	0
Piapoco (7)	0	0	0.286
Surui (24)	0	0	0
Teribe (3)	0	0.333	0
Ticuna (6)	0.167	0	0.167
Toba (4)	0.125	0.125	0.250
Waunana (3)	0	0.167	0.500
Wayuu (11)	0.056	0	0.056
Wichi (5)	0.200	0	0.300
Yaghan (4)	0.125	0.125	0.250
Total (106)	**0.048***	**0.053***	**0.142***

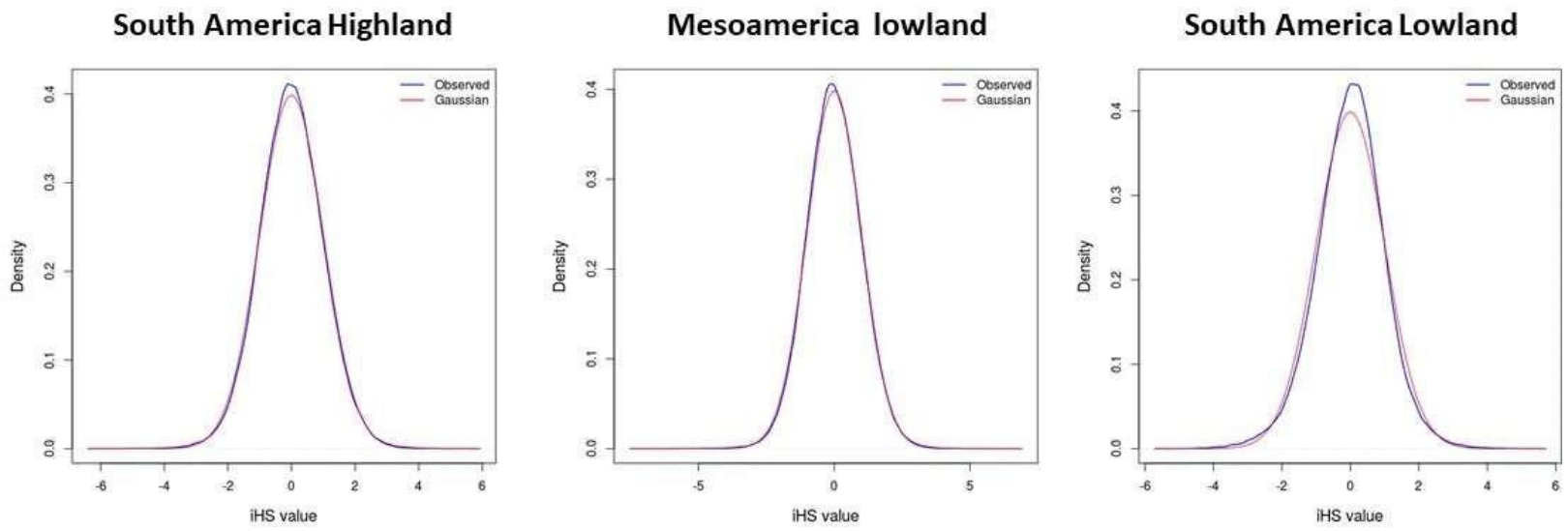


Figure S1. iHS value distribution patterns for all three groups (South American highland, South America Lowland and Mesoamerica Lowland).

	Mean (lowland)	Median (lowland)	95% c.i.	Mean (highland)
SP100 (rs13411586 C)	0.0656	0.0652	(0.0481; 0.0854)	0.3968
SP100 (rs9678342 C)	0.0637	0.0637	(0.0466; 0.0838)	0.3809
SP100 (rs7582700 T)	0.0965	0.0963	(0.0745; 0.1196)	0.3889
DUOX2 (rs269866 G)	0.0714	0.0714	(0.0528; 0.0916)	0.4206
CLC (rs440191 A)	0.1332	0.1335	(0.1071; 0.1599)	0.4524

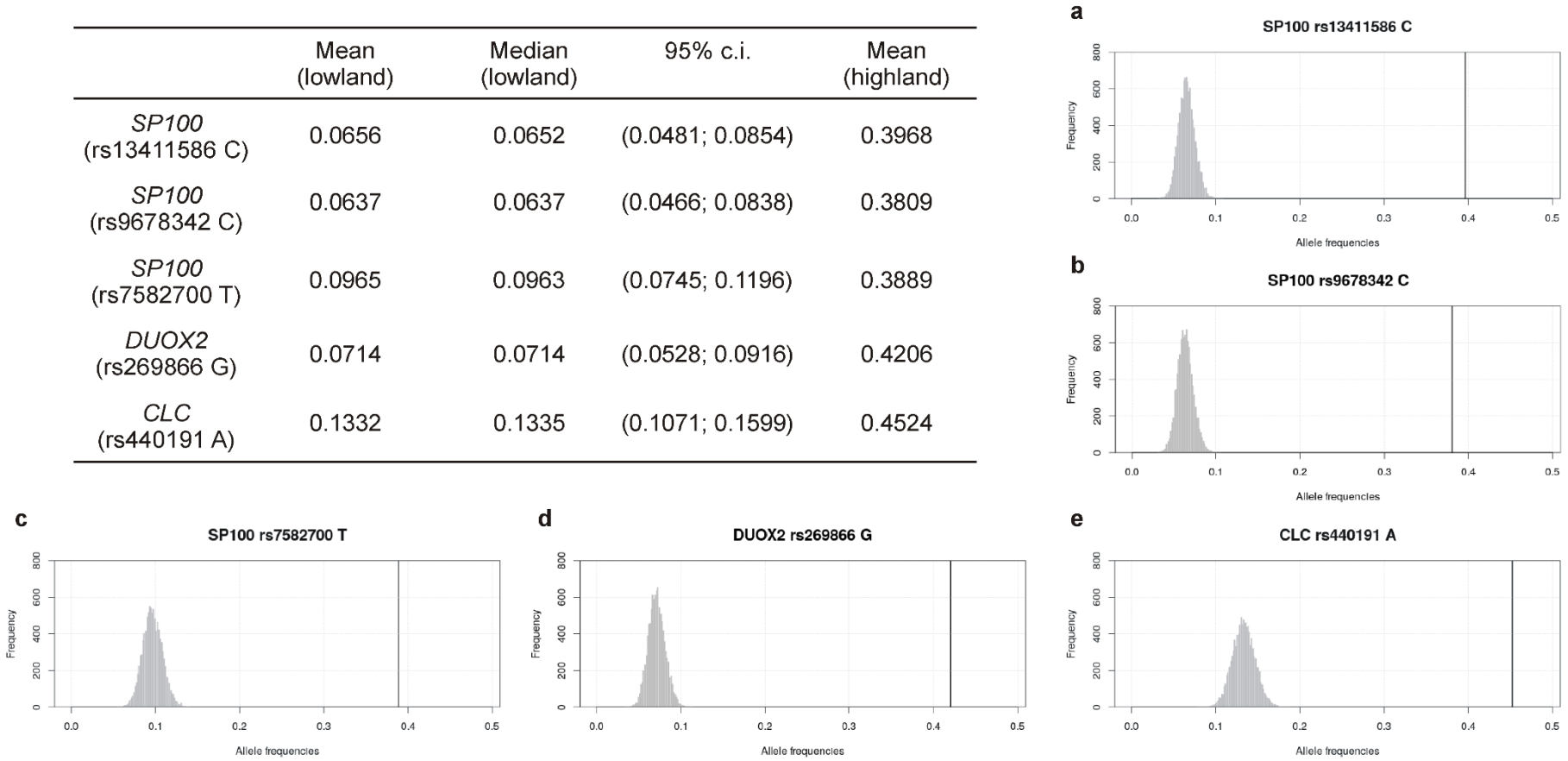


Figure S2. **Bootstrap simulations.** (A) Table showing average values (mean and median) of the allele frequencies for each SNP in lowland populations, as well as, their 95% confidence intervals obtained by simulation, and the average allele frequency of the candidate variant in highland populations. (B–F). Distribution of allele frequencies obtained by 10,000 simulations for lowlanders considering all markers in putative selection. The corresponding average allele frequencies observed for highland populations are represented by black vertical lines.

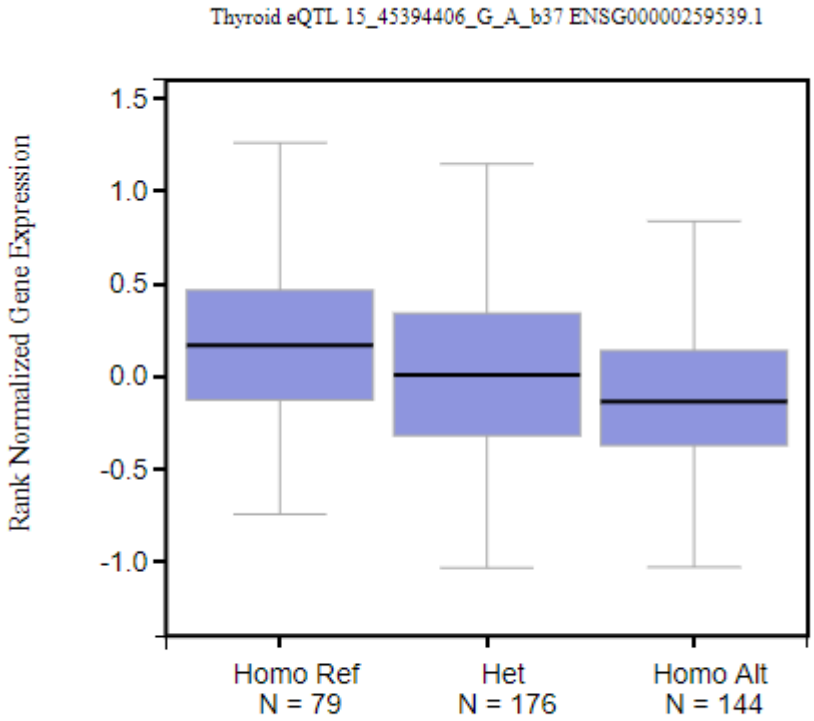


Figure S3. Differential expression of the DUOX2 putatively selected allele (rs269866 G) for a differential effect in gene expression in the thyroid tissue (https://www.gtexportal.org/home/, accessed 26/03/2018).

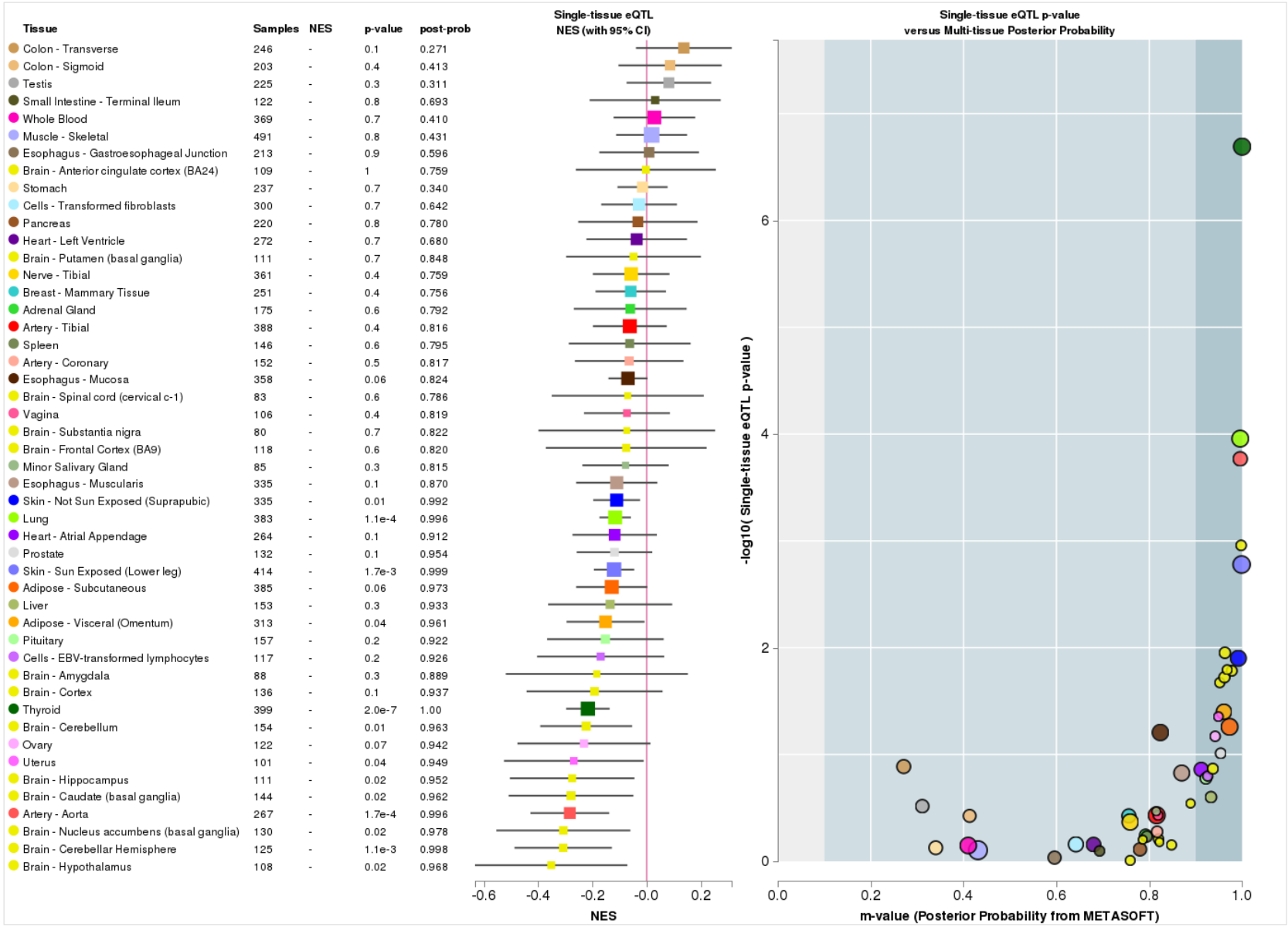


Figure S4. Posterior probabilities of DUOX2 putatively selected allele (rs269866 G) for a differential effect in gene expression in multiple tissues (https://www.gtexportal.org/home/, accessed 26/03/2018).

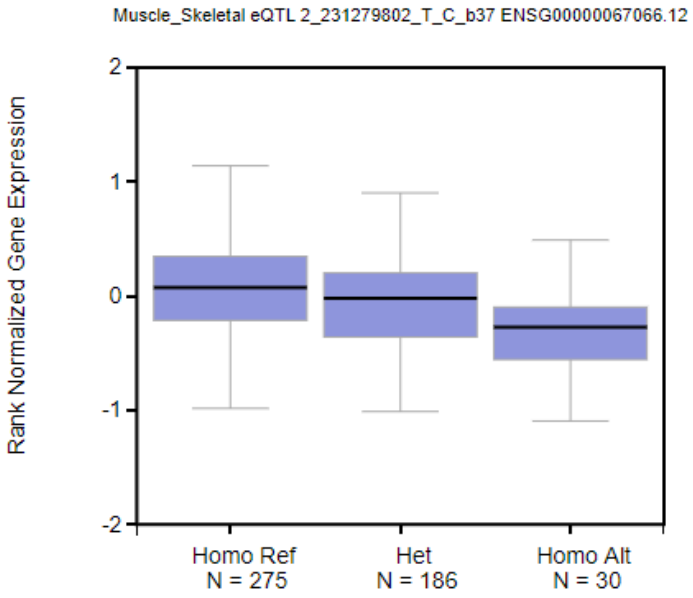


Figure S5. Differential expression of the SP100 putatively selected allele (rs13411586 C) in the skeletal muscle (https://www.gtexportal.org/home/, accessed 26/03/2018).

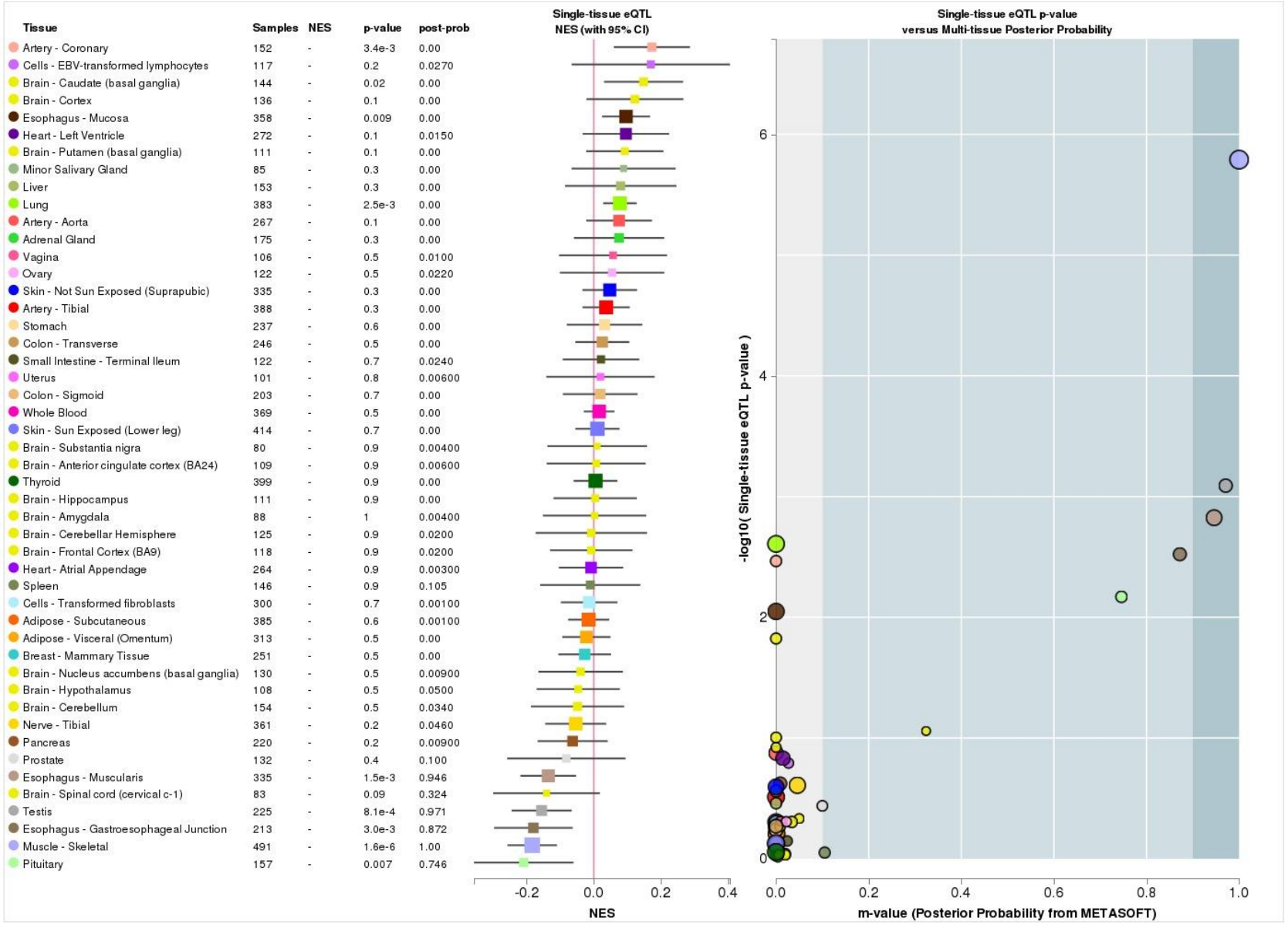


Figure S6. Posterior probabilities of SP100 putatively selected allele (rs13411586 C) for a differential effect in gene expression in multiple tissues (https://www.gtexportal.org/home/, accessed 26/03/2018).

CAPÍTULO 5

Manuscrito em preparação

Andean high altitude as a potential selection factor influencing haplotype*HLA-G* 3'UTR distributions in South American Natives

Vanessa Cristina Jacovas, Rafael Tomoya Michita*, Rafael Bisso-Machado,Guillermo Reales, José Raul Sandoval, Alberto Salazar-Granara, Francisco Mauro Salzano*, José Artur Bogo Chies and Maria Cátira Bortolini.

* *In memoriam*.

Andean high altitude as a potential selection factor influencing haplotype*HLA-G* 3'UTR distributions in South American Natives

Vanessa Cristina Jacovas[1]*, Rafael Tomoya Michita[1,2]*, Rafael Bisso-Machado[1], Guillermo Reales[1], José Raul Sandoval[3], Alberto Salazar-Granara[3], Francisco Mauro Salzano**, José Artur Bogo Chies[1] and Maria Cátira Bortolini[1]

[1] Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.
[2] Human Molecular Genetics Laboratory, Universidade Luterana do Brasil (ULBRA), Canoas, Rio Grande do Sul, Brazil.
[3]Facultad de Medicina Humana, Universidad de San Martin de Porres (USMP), Lima, Peru.

* These authors contributed equally to this work.
** *In memoriam*.

Corresponding authors: Maria Cátira Bortolini and José Artur Bogo Chies, Department of Genetics, Biosciences Institute, Universidade Federal do Rio Grande do Sul, Bento Gonçalves Avenue 9500, Campus do Vale, CEP 91501970, Porto Alegre, RS, Brazil.
Phone: +55 51 3316-6740 and +55 513308-9844; Fax: +55 51 3316-7311;
E-mails: maria.bortolini@ufrgs.br and jabchies@terra.com.br.

Abstract

Physiological response to hypoxia involves a wide range of biological mechanisms. HLA-G plays a fundamental role in several biological processes, including human reproduction and hypoxia response, thus being a potential candidate gene involved in human high-altitude adaptation. The genetic variation in the promoter region of *HLA-G* is reported to influence HLA-G expression patterns. A large number of studies show that hypoxia is a condition that modulates the production of HLA-G in both micro and macro environment contexts. However, the current knowledge of *HLA-G* genetic variation in Native American populations, and especially at the 3' untranslated region (UTR) is still unexplored. In the present study, we aimed to identify whether combinations of the functional *HLA-G* 3'UTR variants have been evolutionarily co-opted for human high-altitude adaptation in the Andes. For this purpose, we evaluated the *HLA-G* 3'UTR region of 301 South Native Americans from 17 populations located in Andes area (Highland; >2,500m) and Lowland region (<2,500m) of South America. In total, 11 haplotypes were observed. The most frequent haplotypes (UTR-1, UTR-2, and UTR-3) previously associated with high and intermediate production of HLA-G account for ~76% of haplotypes found in Lowland, while in Highland population it reaches up to 92.3%. The ancestral UTR-5 haplotype, which is associated with low HLA-G levels, reaches up to 21.6% and 5.1% in Lowland and Highland, respectively. The correlation analyses showed that UTR-2 frequency positively correlates with altitude, while it decreases according to the oxygen concentration and temperature, an opposite situation that occurs with UTR-5. We also observed that 31% percent of the polymorphic sites located at *HLA-G* 3'UTR are trans-species, shared by modern and archaic humans, suggesting that combinations of ancient and more derived alleles (and their combinations, *i.e.* haplotypes) may have been co-opted for relatively recent adaptations such as those experienced by modern humans in the Andes. On the other hand, this same genetic repertoire can be unfavorable to Andeans in the present day, considering the tumor microenviromental context, since it provides tools for the cancer cell to survive and multiply.

Keywords: Adaptation; high-altitude; HLA-G 3'UTR; Native Americans; Andeans.

Introduction

The study of human adaptation in the Andes has attracted much interest by researchers since the beginning of the 20th century, still, in 1927, a Peruvian physician named Carlos Monge Medrano, who later specialized in altitude medicine, organized an expedition to the Peruvian highlands. There, he confirmed the existence of physiological mechanisms that over the centuries acclimated Highland populations to live in low oxygen concentrations (hypoxia), and other harsh conditions of high altitudes (Monge et al., 1948). Noteworthy, that is relatively well-known that plastic physiological and functional response (*i.e.,* increase in pulmonary ventilation and red blood cell count to raise hemoglobin concentration and minimize the effects of hypoxia) occurs when individuals from lowlands visit or reside for a short-time at highlands (Moore, 2017). These changes are transient, *i.e.,* reversible and do not usually cause damage, except individuals who suffer some form of altitude sickness, which typically occurs only above 2.500 meters. The most common symptoms include headache, loss of appetite, insomnia, and nausea, but the severe forms may lead to death (Meier et al., 2017; reviewed in Simancas-Racines et al., 2018 and references therein).

It is known that these types of short-time or transitory biological "adaptations" are different from those observed in long-time native residents. In the temporal depth of an evolutionary context several morphological and physiological traits are required for life-long living at high-altitude, including adaptations for permanent hypoxia, high ultraviolet radiation (UV) and cold temperatures (Beall et al., 2007). Thus, natural selection is expected to leave footprints in the genome of the individuals inhabiting these regions with extreme climatic conditions such as Andean Altiplano; a place inhabited continuously since ~12,000 years ago (Rademaker et al., 2014).

The Hypoxia-Inducible Factor (HIF) pathway is the master transcriptional regulator of hypoxia response in metazoans (Bigham and Lee, 2014; Bighman, 2016). The HIF gene products respond to insult promoted by hypoxia, inducing various mechanisms of tolerance, such as red blood cell production and angiogenesis (Haase, 2013; Zonneveld et al., 2019). In a tumor microenvironment, where hypoxia is frequently observed, the cancer cell survival and propagation dependes on its hability of induce these mechanisms (Zonneveld et al., 2019). Interestingly, genes belonging to the HIF pathway, such as *EDN1, EDNRA, EGLN1, NOS1,* and *VEGFA* appear to have been essential for human

adaptation in the three distant geographic regions (Andes, Tibet, and Ethiopia; Bighman, 2016), but with similar climatic conditions related to the high altitudes, a "macro environment" with permanent hypoxia. *HIF1α* and *MDM2*, in turn, seems to have been evolutionarily co-opted in Ethiopia and Tibet, and Tibet and Andes, respectively (see other examples in Bighman, 2016).

Andeans, Tibetans, and Ethiopians present also particularities regarding their selected genetic repertoires potentially involved in high-altitude adaptation. For example, we reported a previously unexplored route, describing that a particular combination of alleles of the *TP53* pathway genes, which regulate the cell cycle, DNA repair, senescence, apoptosis, and reproductive success has been evolutionarily co-opted for adaptation to the Andean environmental (Jacovas et al., 2015; 2018). Hitherto, similar set of common variants of the *TP53* pathway genes (for instance, *USP7, LIF, MDM2, SP100, DUOX2* and *CLC*) have not been correlated to this same context in Ethiopians and Tibetans (Jacovas et al., 2015, 2018). Interestingly, non-synonymous changes in the orthologue *TP53* were described as having an essential role in high-altitude adaptation in Tibetan rodent species (Zhao et al., 2013). These findings exemplify the existing differences in the genetic repertoire among species or even within single species but also show similar functional effects on selected phenotype.

HLA-G is a non-classical MHC class-I molecule with immunoregulatory properties, and depending on the biological context it may activate or induce immunologic tolerance by inhibiting different cells involved in innate and adaptive immunity (Sabbagh et al., 2014). HLA-G is translated in different isoforms (soluble and membrane-bound) yielded by alternative splicing of mRNA, and are abundantly expressed on extravillous trophoblast cells. HLA-G is an important component for extravillous trophoblast invasion to form the placenta in pregnancy, and fundamental to the establishment of immune tolerance at the maternal-fetal interface and blastocyst implantation (Mouillot et al., 2007; Ferreira et al., 2017; Garziera et al., 2017; Reeves and James, 2017), normal conditions that are important to ensure reproductive success.

Interestingly, these gestational processes occur under hypoxia, a critical factor in stimulating HLA-G synthesis (Mouillot et al., 2007; Garziera et al., 2017). Noteworthy, soluble HLA-G (sHLA-G) levels are associated with tumor progression, a condition where hypoxia takes place (Nagamatsu et al., 2004; Yaghi et al., 2016). Moreover, the expression

of HLA-G on cancer cells, associated with hypoxia, is considered a critical immune checkpoint exploited by the tumor angiogenesis and escape from host immune surveillance (Garziera et al., 2017). Then, hypoxia induces HLA-G expression in both normal and pathological microenvironments and might have harmful or beneficial effects, depending on the context (Alegre et al., 2014).

Several mechanisms and elements related to the regulation of *HLA-G* have been described. For example, experimental procedures revealed the role of the Hypoxia-Inducible Factor-1 (HIF-1), which binds to the hypoxia-responsive elements located at the coding and promoter region of *HLA-G* (Yaghi et al., 2016; Garziera, et al., 2017). Despite being a short region (~358bp), many single nucleotide polymorphisms (SNPs) in *HLA-G* 3'UTR have been reported in human populations (see Table S1; Castelli et al., 2011; Sabbagh et al., 2014). Some of them seem to influence HLA-G expression through atypical mechanisms compared to others found in the HLA class I genes (Castelli et al., 2010; Porto et al., 2015). Then, since they are functional, polymorphic sites located at *HLA-G* 3'UTR have been associated with risk for malignancies, autoimmune disorders, infectious diseases, recurrent pregnancy loss, among other conditions (Medeiros et al., 2018; de Almeida et al., 2018; Craenmehr et al., 2019).

A high linkage disequilibrium (LD) is observed among *HLA-G* 3'UTR variant alleles, suggesting that haplotypes should be evaluated due to the potential cumulative effects of alleles in HLA-G expression patterns (Alvarez et al., 2009; Castelli et al., 2010; Sabbagh et al., 2014). The haplotype named UTR-1 is usually associated with highest sHLA-G levels since it comprises the variants independently associated with high HLA-G production. On the other hand, UTR-5 (putative ancestral; Sabbagh et al., 2014) and UTR-7 are associated with lower sHLA-G levels, whereas UTR-2, 3, 4 and 6 exhibit intermediate levels (Martelli-Palomino et al., 2013).

Different from other major human geographical groups (Sabbagh et al., 2014), Native Americans are still uncharacterized in terms of *HLA-G* 3'UTR diversity. In the present study we aimed to answer the following questions: 1) HLA-G has been evolutionarily co-opted for human long-time high-altitude adaptation in the Andes? 2) The functional *HLA-G* 3'UTR haplotypes are relevant in this evolutionary scenario? In order to answer these questions, *HLA-G* 3'UTR was sequenced in a comprehensive set of Native American populations covering highland (>2,500m) and lowland (<2,500m) ecoregions

from South America. Additionally, the genetic variables were correlated with climate- and environmental-related variables, such as altitude, temperature variation, O_2 concentration, and UV radiation index.

Material and Methods

Samples and ethical procedures

DNA samples were obtained from 301 volunteers characterized as Native American or as having large Native American ancestry (>90%; Sandoval et al., 2013). Individuals are from 17 populations located in two different South American ecoregions, based on altitude: Highland populations (≥2,500m) and Lowland populations (<2,500m). Highland populations include Uros (n=21), Aymara (n=6), Quechua (n=15), Anapia (n=15), Amantani (n=28), Cabanaconde (n=16), Chivay (n=17), Taquile (n=43), and Yanque (n=10). These populations are located in the Andean region, including the Lake Titicaca islands and their vicinity. Excepted by the Uros, they can be classified as agro-pastoralists, since they adopted practices of cultivation and grazing of camelids thousands of years before the arrival of the European colonizers. Lowland populations encompasses Andoas (n=61), Guarani Ñandeva (n=6), Guarani Kaiowá (n=6), Xavante (n=20), Xikrin (n=6), Cinta-Larga (n=7), Zoró (n=4), Lengua (n=11). These populations are characteristic of the Amazonian tropical forest, Brazilian central plateau and Chaco, and can be classified as basically hunter-gatherers, since up to the present day, or even a few decades ago, they maintained the habits of life and diet characteristic of these groups.Additional information about the populations investigated here see Sandoval et al., (2016) and Reales et al., (2017).

Ethical approval for our study was provided by the Brazilian National Ethics Commission (CONEP Resolution no. 123/98) and by ethics committees in the countries where non-Brazilian samples were collected. All sampling procedures were performed according to the Helsinki Declaration. The ethics committees also approved the oral consent procedure - since most participants were illiterate - as well as the use of those samples for population and evolutionary studies.

HLA-G 3'UTR Sequencing

Genomic DNA was obtained from plasma, glycerolized red blood cells or whole blood using the QIAamp DNA MiniKit (Qiagen, Hilden, Germany). The following primers were used to sequence the *HLA-G* 3′ untranslated region (~358bp): Forward: 5'-TGTGAAACAGCTGCCCTGTGT-3'; Reverse (5' to 3'): 5'-CTGGTGGGACAAGGTTCTACTG-3′ (Castelli et al., 2010; Michita et al., 2016). Polymerase chain reaction (PCR) conditions were as follows: first step (initial denaturation): 94°C for 5 min; second step (32 cycles): 94°C for 30 s, 65.5°C for 30 s and 72 °C for 1 min; third step (final extension): 72°C for 5 min. The PCR products were directly sequenced using the reverse primer in an ABI 3730 XL DNA sequencer (Applied Biosystems, Foster City, CA, USA). Sequences were aligned and their qualities, as well as the assessment of the accuracy of the resulting data, were ascertained using the FinchTV software version 1.4.0 (free available at http://www.geospiza.com/Products/finchtv.shtml). The complete list of all SNPs evaluated in this region is shown in **Table S1**. *HLA-G* 3'UTR haplotypes were named according to previous studies (Alvarez et al., 2009, Castelli et al., 2010 and Sabbagh et al., 2014).

Data analyses

Analysis of molecular variance (Weir and Cockerham, 1984; Excoffier et al., 1992; Weir, 1996) was evaluated by Arlequin software using F_{ST} (among-population component of variance) and F_{CT} (between-group component of variance) statistics. All analyses were performed grouping the populations in highlanders - individuals residing permanently above sea level (>2,500m; Moore, 2001) - and lowlanders (those who reside permanently <2,500m). For both hierarchical levels, a *p*value<0.05 was considered statistically significant.

To investigate possible adaptations to climate conditions in highlands and lowlands, we performed Spearman's correlation tests between haplotype frequencies and ten climatic and environmental variables (**Table S2**) using SPSS v.18.0 for Windows (SPSS Inc., Chicago, Illinois, USA). Environmental data were collected for each population using the SoDa Service for irradiation data, WorldClim for bioclimatic data and altitude.org for barometric pressure (http://www.soda-is.com; http://www.worldclim.org; http://www.altitude.org/air_pressure.php; last access: May 31th, 2017). The map indicating

the original populations studied was generated using DIVA-GIS (**Figure 1**; version 7.5.0.0; http://www.diva-gis.org/; Hijmans et al., 2001).

Linkage disequilibrium (LD) was evaluated by r^2 and D' coefficients using the Haploview 4.1 software (Barrett et al., 2005). Haplotypes were inferred using a Bayesian method implemented in the PHASE software version 2.1 using default parameters (Stephens et al., 2001). Haplotype networks (Bandelt et al., 1999) were constructed using the Network software (version 5; http://www.fluxus-engineering.com). Possible reticulations were resolved by maximum parsimony calculation (Polzin and Daneschmand, 2003).

We performed the Ewens–Watterson test (Ewens 1972; Watterson, 1978) based on haplotypes frequency distributions to detect departures from selective neutrality, using Arlequin 3.5.1.2 (Excoffier and Lischer, 2010). Ewens–Watterson test can potentially reveal whether demographic events (for example, rapid population growth) and directional (observed homozygosity higher than expected homozygosity) and/or balancing (observed homozygosity lesser than expected homozygosity, indicating an excess of heterozygosity) selection are acting on a particular locus across populations. A p-value <0.05 was set to detect signals of balancing selection (or recent bottleneck) and a p-value >0.95 was set to detect signals of positive selection (or population growth) (McClelland et al., 2013).

Results

Figure 1 shows the original geographic location of the populations included in the present study. Based on the 1k Genome Project, Sabbagh et al., (2014) observed 19 variable sites at *HLA-G* 3′UTR region, giving substantial support to previous studies with smaller and less comprehensive samples (Tan et al., 2005; Castelli et al., 2011). Ten of them are monomorphic in our sample, while nine show polymorphic variations (**Table S3**).

With these 19 *loci*, several haplotypes can be inferred. The most frequent in both lowland and highland groups is UTR-2, followed by UTR-1 and UTR-3. Noteworthy, UTR-1, associated with higher levels of sHLA-G, reaches high distributions around the world, regardless of its derived state (four mutational steps from the ancestral haplotype UTR-5; **Table 1**, **Figure 2**) relative to the others, UTR-2, UTR-3 (both with intermediate levels of sHLA-G). The sum of these most frequent haplotypes (UTR-1, UTR-2, and UTR-3) represent about 70% of haplotypes found in Africa, 71% in Europe and 81% in Asia

(Sabbagh et al., 2014; **Table S4**). In admixed populations from America, the number is 70.1% (Sabbagh et al., 2014). The number found in South America lowlands, 76%, is intermediate than those found in the other continents, whereas in highlands is higher, 92.3%.

On the other hand, UTR-5 considered the ancestral haplotype (Sabbagh et al., 2014) and associated with lower sHLA-G levels (Martelli-Palomino et al., 2013) has a higher frequency in the lowlands (mean = 21.6%) when compared to highlands (5.1%). The last number is intermediate in comparison to those found in other continental groups, but the former is impressively larger (Africa, 8.1%; Europe, 2.7%; Asia, 1.4%; Sabbagh et al., 2014).

Interestingly, these compelling numbers, relative to the frequencies of the ancestral UTR-5 (21.6% and 5.1%) suggest general trends since the differences in the distributions within each group are not enough to indicate significant structuring, measured by F_{ST} (**Table 1**). In other words, lowland populations tend to present high UTR-5 frequencies, while highland show low UTR-5 frequencies. Moreover, between-group (F_{CT}) component of variance is >11% (p<0.001), indicating that lowland and highland groups differ significantly regarding their UTR-5 distributions, a pattern not seen in other haplotypes and something surprising, considering the low differentiation found even between continental populations (Sabbagh et al., 2014).

Some haplotypes were exclusively found in highlanders, but with low or intermediate frequencies: the UTR-8 (Amantani, 1.8%), UTR-18 (Quechua, 3.3%; Yanke, 5.6%) and UTR-30 (Taquile, 1.2%; Uro, 2.4%). In contrast, UTR-9 (Guarani Kaiowa, 8.3%; Lengua, 2.5%), UTR-13 (Cinta Larga, 14.3%) and UTR-21 (Xavante, 2.5%) were exclusively found in lowland populations.

The haplotypes observed in the present study were correlated with ten environmental variables (**Table 2**; **Table S5**). Statistical correlations were observed among UTR-2 and UTR-5, and relevant climatic conditions in the context of high and low altitudes, respectively. UTR-2 and UTR-5 frequencies show moderate but significant correlation with altitude (rho = 0.56 and -0.60; p = 0.02 and 0.01, respectively), O_2 concentration mmHg (rho = -0.567 and 0.58, p = 0.014 and 0.01), and annual mean temperature (rho = -0.534 and 0.59, p = 0.02 and 0.01). These results indicate that UTR-2 frequency tends to increase with altitude, and with low concentration of oxygen and

temperature, an opposite situation that occurs with UTR-5. On the other hand, UV radiance has a moderate and negative correlation with UTR-5 distribution (rho = -0.52, p = 0.03), whereas a moderate negative correlation (rho = -0.56, p =0.02) was identified between UTR-2 distribution and isothermality, a ratio of the mean diurnal range to the annual temperature range. Notwithstanding, these findings and general tendencies, the frequencies of UTR-2 within each group floats in a significant way, to the point of generating structuring as measured by F_{ST} (7.8%, p = 0.007 and 11.3%, p = 0.013, for highland and lowland, respectively; **Table 1**).

Figure 2 shows a median-joining haplotype network with a multimodal pattern. The high frequencies of some haplotypes (UTR-1, 2, 3 and 5 for lowlands and UTR-1, 2, 3 for highlands), concomitant with presence of "peripheral" variants with intermediate and low frequencies, even in the face of loss due to drift, indicate balancing selection, a microevolutionary phenomenon often evoked to explain the diversity patterns of HLA *loci* (Nielsen et al., 2009; Viscardi et al., 2018). However, the Ewens-Watterson test does not show significant values (p< 0.05 or p> 0.095) for both groups, indicating no deviation from the neutral expectations (mean Watterson F´s p-values = 0.49110 and 0.54814 for highland and lowland, respectively).

Interestingly, we observed that for six *HLA-G* 3'UTR *loci* (31% of the total) two segregating alleles are seen in archaic humans: +3111 (rs554784083) is a potential polymorphic locus in Neanderthal, whereas +3010 (rs1710), +3027 (rs17179101), +3035 (rs17179108), +3142 (rs1063320), and +3187 (rs93801420) in Denisova (**Table S1**). Only *locus* +3111 (rs554784083) is not polymorphic in the Native American populations investigated here. One of the Neanderthal haplotype inferred from data present in **Table S2**, is the UTR-6 (associated with intermediate sHLA-G levels; Martelli-Palomino et al., 2013), curiously more common in Africa (mean of 8.1%) than in Europe (2.9%), the later a continent where the Neanderthals lived. The other Neanderthal haplotype, with the allele A at position +3111, for our knowledge, has not been described in modern human populations (Castelli et al., 2011; Sabbagh et al., 2014). For Denisova, only one genome sequence is avaiable, making it impossible to infer the haplotypes. However, this individual has many loci in heterozygosis, indicating that he lived in a more diverse population relative to Neanderthal, corroborating our previous studies (Paixão-Cortes et al., 2013).

This kind of allele sharing between modern and archaic humans can occur for two reasons: introgression or a long-term maintenance of trans-species polymorphisms (trans-SNPs), *i.e.* ancient polymorphisms that survived in derived taxa, a classic indicator of balancing selection in the major histocompatibility complex (MHC) (Takahata and Nei, 1990; Clark, 1997; Grimsley et al., 1998; Ségurel et al., 2013; Azevedo et al., 2015).

Discussion

The expression of HLA-G induced by hypoxia in normal and pathological microenvironments is well-documented in abundant bibliography (Mouillot et al., 2007; Nagamatsu et al., 2004; Alegre et al., 2014; Yaghi et al., 2016; Garziera et al., 2017). However, considering macro environment contexts (external conditions such as those found at high altitudes) we were able to find only one study. Bourguignon et al. (2010) reported the hypoxia in high altitude as an important factor in the regulation of HLA-G expression, at least considering transient circumstances. In this study they evaluated the peripheral blood of climbers before departure of the expedition and during their ascent to and descent from summit of Mount Everest (Himalaya, Tibet). The authors showed an increase in sHLA-G levels in four of six climbers during the ascent on Mount Everest. In the same study, four Tibetan Sherpas (native people of the Himalaya) were also investigated. Sherpas' sHLA-G levels were slightly lower than those of mountaineers, with one exception. Unfortunately, blood sampling was not performed for Sherpas at the highest altitudes (> 7,800 m), thus not permitting the evaluation of sHLA-G amount variations in the same conditions of the non-Tibetans (Bourguignon et al., 2010). The authors did not perform any genotyping of these volunteers, so their genetic repertoire, relative to candidate genes such as *HIF*-pathway genes and *HLA-G* are unknown.

Our data show high frequencies in Andean populations of the derived haplotypes (UTR-1, UTR-2, and UTR-3) that would produce intermediate to high levels of sHLA-G. Interestingly, UTR-2 is significantly correlated with conditions found at high altitudes of the Andes. Multiple alleles/haplotypes maintained in a population over relatively long evolutionary periods are characteristic of balancing selection. Additionally, it is also observed several alleles/haplotypes with low/intermediate frequencies (Siewert and Voight, 2017). **Table 1** and **Figure 1** illustrate this situation in both groups, where the haplotypes UTR-1, UTR-2, and UTR-3 can be assumed as the balanced variants.

Noteworthy, UTR-5 seems to have become a balanced variant in the lowlands, while remaining as a peripheral haplotype in the Andes.

Our results also revealed that UTR-2 is significantly correlated with low O_2 concentration and cold temperature, whereas UTR-5 is correlated with conditions found in lowlands, including low UV incidence. Haplotype UTR-5 may not be favorable in hypoxic macro environments since it usually associates with low sHLA-G production in opposite to UTR-2.

Recently, we investigate the role of balancing selection (overdominance/ heterozygote advantage) in human evolution using available exomes from southern Saharan Africans representing the *Homo sapiens* (to minimize the possibility of sharing alleles by introgression) and archaic humans (*Homo neanderthalensis* and Denisova specimen). We identified a significant excess of trans-SNPs (N=547) in 1,754 genes of the immune system in comparison to the background genomic distribution of trans-SNPs generated with a random gene set, suggesting heterozygote advantage in maintaining these polymorphisms (balancing selection). Noteworthy, that of the 547 trans-SNPs identified by us (Viscardi et al., 2018), eight (rs1630223, rs1630185, rs1130355, rs1130356, rs375611189, rs79264452, rs74547057, rs1130363) are located at coding regions of *HLA-G* (Vircardi et al., 2018).

Here, we describe for the first time that the 3'UTR region of *HLA-G* presents also an important proportion of polymorphisms shared by modern and archaic humans (+3111, rs554784083 in Neanderthal; +3010 rs1710, +3027 rs17179101, +3035 rs17179108, +3142 rs1063320, and +3187 rs93801420 in Denisova). Previous studies with *EPAS1* (*HIF-2α*) indicated in Tibetans advantageous alleles for life at high altitude originating through ancient admixture with Denisovans (Huerta-Sánchez et al., 2014). Qin and Stoneking (2015) found a widespread signal of a very low level of Denisovan ancestry across Eastern Eurasian and NativeAmerican populations. Then, a similar case of that described for the *EPAS1* can not be completely ruled out to explain the shared polymorphisms between Denisovans and Andeans found in present study. However, it is more reasonable to assume that these shared polymorphisms are trans-SNPs, maintained for thousands of years by balanced selection, since this phenomenon has recurrently been described when immune system genes are investigated (Viscardi et al., 2018).

Earlier we suggested that trans-SNPs of the immune system could have played an essential immunological role during speciation and migration of the genus *Homo* in a probable similar context to their hominin common ancestral (Viscardi et al., 2018). Several studies, including with trans-SNPs, have revealed that individuals heterozygous for genes with critical roles in the immune system are more effective in the defense against pathogens, while at the same time presenting only a moderate inflammatory response (Cagliani et al., 2008; Leffler et al., 2013; Azevedo et al., 2015; Teixeira et al., 2015). Another example of the delicate balance regarding the immune system is the competing needs of the fetoplacental unit to inactivate maternal immune cells to ensure its survival, and at the same time to maintain a sufficient number of immune competent maternal cells in the uterus to ward off infection (Tan et al., 2005). Similarly, it is tempting to speculate that at least in the Andes, there is a delicate balance involving evolutionary forces maintaining alleles/haplotypes that promote high (UTR-1) and intermediate (UTR-2 and UTR-3) production of HLA-G, while maintaining in relatively low frequencies others, such as ancestral UTR-5 (related the low HLA-G production).

Despite this pieces of evidence, as well as signal of the balancing selection in the *HLA-G* 3'UTR region obtained by other authors considering other human populations (Castelli et al., 2011), the neutrality Ewens-Watterson test used here was not able to discard stochastic and demographic events as the standard causes of diversity pattern found in both lowland and highland groups (*i.e.*, other classic tests, such as Tajima's D and Fu's *Fs* tests also do not show significant values; data not shown). Fijarczyk and Babik (2015) highlight that balancing selection may be a transient state, leaving marks so subtle that their detection may be difficult using current tests, leading to a large number of false negatives. It is also worthy to mention that Native Americans have a very complex demographic history, which is an additional confounding factor in the analyzes since it can mimic or camouflage the action of natural selection.

No single mechanism could account alone for an effective adaptive response for a long-life of humans at high altitudes (Beall 2014). Here, we provide pieces of evidence that suggests that HLA-G level, mediated by functional sites at promoter *HLA-G* 3'UTR, may be an additional element of the genetic repertoire behind the long-life adaptation to high altitudes in the Andes. This suggestion leads to an affirmative answer to the two questions initially formulated. Additionally, it is possible to go beyond and raise other instigating

issues: some of the functional sites/haplotypes may be being maintained by at least 400-275 thousand years (*i.e*., the split between *Homo* sapiens and Neanderthal-Denisova clade; Endicott et al. 2010; Prüfer et al. 2014) suggesting that a delicate set of ancient and derived alleles (and their combinations, *i.e.* haplotypes) may have been co-opted for relatively recent adaptations (~12,000 years) such as those experienced by modern humans in the Andes.

Interestingly, government surveys in Peru indicate that the rate of gestational and postpartum complications in Aymara Andean regions is lower than the national average (see Jacovas et al, 2015 and citations therein). On the other hand, the incidence of cancer in highlanders, including Andeans, is higher (Volskarides et al., 2018). Considering these informations, the role of HLA-G, as well as our findings, a scenario may be suggested: a complex and multifunctional genetic repertoire (*e.g HIF* and *TP53* patways; Bighman, 2016; Jacovas et al., 2015, 2018), which involves also a delicate combination of ancient and derived HLA-G alleles/haplotypes (probably maintained by balancing selection), allowed tolerates hypoxia at the macro environment level (life at high atitude), with possible positive reflexes also in the healthy gestation microenviromental level. As a result, the Andeans are well adapted to a long-life at high altitudes. On the other hand, this same genetic repertoire can be unfavorable to these individuals in the present days, considering a specific patological microenviromental context, since it provides tools for the cancer cell to survive and multiply (from a clinical point of view, hipoxia of tumor is associated with poor outcome; Zonnevel et al., 2019). Furthermore, even considering a supposed more efficient system of repair and/or programmed cell death (apoptosis) in highlanders, mediated by p53 (Jacovas et al., 2015) required to avoid severe damage to DNA and/or cells due to high incidence of UV and cold temperatures, it is less efficient to control the incidence and progression (associated also with many somatic tumor mutations subsequent to mutation of origin; Petljak et al., 2019) of cancer at least relative to lowlanders. This scenario makes sense considering that the vast majority of the human cancer cases emerge after the reproductive period, making the disease, as it is know today, irrelevant in an evolutionary context.

In addition, the implication of the homogeneously high frequency of UTR-5 in the lowland populations still needs to be further evaluated, since it is a striking contrast with

the general diversity pattern found in South American Native hunter-gatherers when neutral *loci* are considered, and significant genetic structuring is a rule (Wang et al., 2007).

Finally, and given the complexity of the theme, only more genetic and functional studies with Andeans, as well as with Tibetans and Ethiopians, can corroborate our findings and test the hypothesis suggested here regarding the relevance of *HLA-G* in an evolutionary context and its implications for the health/disease paradox of modern populations living at high altitude.

Acknowledgments

The authors are grateful to all individuals who donated the samples analyzed in the study. We thank Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (grant No. 14/2253-2 PqG) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant No. 473115/2011-5 and 305839/2015-2) for funding experimental procedures, as well as fellowships. We also thank Profs. Eduardo M. Tarazona-Santos (UFMG, Brazil) and Maria Luiza Petzl-Erler (UFPR, Brazil) for providing the Peruvian Quechua and Guarani samples, respectively.

References

Alegre E, Rizzo R, Bortolotti D, et al. Some basic aspects of HLA-G biology. J Immunol Res. 2014; 2014:657625.

Alvarez M, Piedade J, Balseiro S, et al. *HLA-G 3′ UTR* SNP and 14-bp deletion polymorphisms in Portuguese and Guinea-Bissau populations. Int J Immunogenet 2009; 36:361-366.

Azevedo L, Serrano C, Amorim A et al. Trans-speciespolymorphisminhumansandthegreat apes is generally maintained by balancing selection that modulates the host immune response. Hum Genomics. 2015; 9:21.

Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999; 16:37-48.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21(2):263-265.

Beall CM. Detecting natural selection in high-altitude human populations. Respir Physiol Neurobiol. 2007; 158(2-3):161-171.

Beall CM. Adaptation to High Altitude: Phenotypes and Genotypes. Annu. Rev. Anthropol. 2014; 43:251-272.

Bigham AW. Genetics of human origin and evolution: high-altitude adaptations. Curr Opin Genet Dev. 2016; 41:8-13.

Bigham AW, Lee FS. Human high-altitude adaptation: forward genetics meets the HIF pathway. Genes Dev. 2014; 28(20): 2189-204.

Bourguignon M, Yaghi L, Flajollet S, et al. Increased soluble human leukocyte antigen-G levels in peripheral blood from climbers on Mount Everest. Hum Immunol. 2010; 71:1105-1108

Cagliani R,Guerini FR, Rubio-Acero R,et al. Long-standing balancing selection in the THBS4 gene: Influence on sex-specific brain expression and gray matter volumes in Alzheimer disease. Hum Mutat. 2013;34:743-753.

Castelli EC, Mendes-Junior CT, Deghaide NH, et al. The genetic structure of 3'untranslated region of the HLA-G gene: polymorphisms and haplotypes. Genes Immun. 2010; 11(2):134-141.

Castelli EC, Mendes-Junior CT, Veiga-Castelli LC, et al. A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. Mol Biol Evol. 2011; 28(11):3069-3086.

Clark AG. Neutral behavior of shared polymorphism. Proc Natl Acad Sci U S A. 1997; 94:7730-7734.

Craenmehr MHC, Nederlof I, Cao M, et al. Increased HLA-G expression in term placenta of women with a history of recurrent miscarriage despite their genetic predisposition to decreased HLA-G levels. Int J Mol Sci. 2019; 20(3).

de Almeida BS, Muniz YCN, Prompt AH, et al. Genetic association between HLA-G 14-bp polymorphism and diseases: A systematic review and meta-analysis. Hum Immunol. 2018; 79(10):724-735.

Endicott P, Ho SYW, Stringer C. Using genetic evidence to evaluate four palaeoanthropological hypotheses for the timing of Neanderthal and modern human origins. J Hum Evol. 2010; 59:87–95.

Ewens WJ. The sampling theory of selectively neutral alleles. Theor Pop Biol. 1972; 3:87–112.

Excoffier L, Lischer HEL. Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res. 2010; 10:564-567.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992; 131:479-491.

Ferreira LMR, Meissner TB, Tilburgs T, Strominger JL. HLA-G: At the Interface of Maternal-Fetal Tolerance.Trends Immunol. 2017; (4):272-286.

Fijarczyk A, Babik W. Detecting balancing selection in genomes: limits and prospects. Mol Ecol. 2015; 24(14):3529-3545.

Garziera M, Scarabel L, Toffoli G. Hypoxic Modulation of HLA-G Expression through the Metabolic Sensor HIF-1 in Human Cancer Cells. J Immunol Res. 2017; 4587520.

Grimsley C, Mather KA, Ober C. HLA-H: A pseudogene with increased variation due to balancing selection at neighboring loci. Mol Biol Evol. 1998; 15:1581-1588.

Haase VV. Regulation of erythropoiesis by hypoxia-inducible factors. Blood Rev. 2013; 27: 41–53.

Hijmans RJ, Guarino L, Cruz M, Rojas E. Computer tools for spatial analysis of plant genetic resources data: 1 DIVA-GIS. Plant Genet Resour Newslett. 2001; 127:15-19.

Huerta-Sánchez E, Jin X, Asan Z, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014; 512:194–197.

Jacovas VC, Rovaris DL, Peréz O, et al. Genetic Variations in the TP53 Pathway in Native Americans Strongly Suggest Adaptation to the High Altitudes of the Andes. PLoS One. 2015; 10:e0137823.

Jacovas VC, Couto-Silva CM, Nunes K, et al. Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. Sci Rep. 2018; 8(1):12733.

Leffler EM, Gao Z, Pfeifer S, et al.Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science. 2013; 339:1578–1582.

Martelli-Palomino G, Pancotto JA, Muniz YC, et al. Polymorphic sites at the 3' untranslated region of the HLA-G gene are associated with differential hla-g soluble levels in the Brazilian and French population. PLoS One. 2013; 8(10):e71742.

McClelland EK, Ming TJ, Tabata A, et al. Patterns of selection and allele diversity of class I and class II major histocompatibility loci across the species range of sockeye salmon (Oncorhynchusnerka). Mol Ecol 2013; 22:4783–4800.

Medeiros FS, Martins AES, Gomes RG, et al. Variation sites at the HLA-G 3' untranslated region confer differential susceptibility to HIV/HPV co-infection and aneuploidy in cervical cell. PLoS One. 2018; 13(10):e0204679.

Meier D, Collet TH, Locatelli I, et al. Does This Patient Have Acute Mountain Sickness?: The Rational Clinical Examination Systematic Review.JAMA. 2017; 318(18):1810-1819.

Michita RT, Zambra FM, Fraga LR, et al. A tug-of-war between tolerance and rejection - New evidence for 3'UTR HLA-G haplotypes influence in recurrent pregnancy loss. Hum Immunol. 2016; 77(10):892-897.

Monge C. Acclimatization in the Andes: Historical Confirmations of 'Climatic Aggression' in the Development of Andean Man. Johns Hopkins Press, 1948.

Moore LG. Human genetic adaptation to high altitude. High Alt Med Biol. 2001; 2(2):257-279.

Moore LG. Human Genetic Adaptation to High Altitudes: Current Status and Future Prospects. Quat Int. 2017; 461:4-13.

Mouillot G, Marcou C, Zidi I, G, et al. Hypoxia modulates HLA-G gene expression in tumor cells. Hum Immunol. 2007; 68(4):277-285.

Nagamatsu T, Fujii T, Yamashita T, et al. Hypoxia does not reduce HLA-G expression on extravillous cytotrophoblasts. J Reprod Immunol. 2004;63(2):85-95.

Nielsen R, Hubisz MJ, Hellmann I, *et al.*Darwinian and demographic forces affecting human protein coding genes. Genome Res. 2009; 19:838–849.

Paixão-Côrtes VR, Viscardi LH, Salzano FM, et al. The cognitive ability of extinct hominins: bringing down the hierarchy using genomic evidences. Am J Hum Biol. 2013; 25:702-705.

Petljak M, Alexandrov LB, Brammeld JS, et al. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell. 2019 ;176:1282-1294.e20.

Polzin T, Daneschmand SV. On Steiner trees and minimum spanning trees in hypergraphs. Operat Res Lett. 2003; 31:12-20.

Porto IO, Mendes-Junior CT, Felício LP, et al. MicroRNAs targeting the immunomodulatory HLA-G gene: a new survey searching for microRNAs with potential to regulate HLA-G. Mol Immunol. 2015; 65(2):230-241.

Prüfer K, Racimo F, Patterson Net al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014; 505:43-49.

Qin P, Stoneking M. Denisovan ancestry in east eurasian and native american populations. Mol Biol Evol. 2015; 2:2665-74.

Rademaker K, Hodgins G, Moore K, et al. Paleoindian settlement of the high-altitude Peruvian Andes. Science. 2014; 346:466-469.

Reales G, Rovaris DL, Jacovas VC, et al. A tale of agriculturalists and hunter-gatherers: Exploring the thrifty genotype hypothesis in native South Americans. Am J Phys Anthropol. 2017; 163(3):591-601.

Reeves E, James E. Tumour and placenta establishment: The importance of antigen processing and presentation. Placenta. 2017; 56:34-39.

Sabbagh A, Luisi P, Castelli EC, et al. Worldwide genetic variation at the 3' untranslated region of the HLA-G gene: balancing selection influencing genetic diversity. Genes Immun. 2014; 15(2):95-106.

Sandoval JR, Lacerda DR, Acosta O, et al. Genographic Project Consortium. The Genetic History of Peruvian Quechua-Lamistas and Chankas: Uniparental DNA Patterns among Autochthonous Amazonian and Andean Populations. Ann Hum Genet. 2016; 80(2):88-101.

Sandoval JR, Salazar-Granara A, Acosta O, et al. Tracing the genomic ancestry of Peruvians reveals a major legacy of pre-Columbian ancestors. J Hum Genet. 2013; 58(9):627-634.

Ségurel L, Thompson EE, Flutre T, et al. The ABO blood group is a trans-species

polymorphism in primates. Proc Natl Acad Sci U S A. 2012; 109(45):18493-8.

Siewert KM, Voight BF. Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. Mol Biol Evol. 2017; 34(11):2996-3005.

Simancas-Racines D, Arevalo-Rodriguez I, Osorio D, Franco JV, Xu Y, Hidalgo R. Interventions for treating acute high altitude illness. Cochrane Database Syst Rev. 2018; 30;6:CD009567.

Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001; 68:978-989.

Takahata N and Nei M. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics. 1990; 124:967-978.

Tan Z, Shon AM, Ober C. Evidence of balancing selection at the HLA-G promoter region. Hum Mol Genet. 2005; 14:3619-3628.

Teixeira JC, De Filippo C, Weihmann A, et al.Long-termbalancing selectioninLAD1maintainsamissensetrans-speciespolymorphisminhumans,chimpanzees, and bonobos. Mol Biol Evol. 2015; 32:1186-1196.

Viscardi LH, Paixão-Côrtes VR, Comas D, et al. Searching for ancient balanced polymorphisms shared between Neanderthals and Modern Humans. Genet Mol Biol. 2018;41(1):67-81.

Voskarides K. Combination of 247 Genome-Wide Association Studies Reveals High Cancer Risk as a Result of Evolutionary Adaptation. Mol Biol Evol. 2018; 35(2):473-485.

Wang S, Lewis CM, Jakobsson M, *et al.*Genetic variation and population structure in Native Americans. PLoS Genet. 2007; 3:e185.

Watterson GA. An analysis of multi-allelic data. Genetics. 1978; 88:171-179.

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38:1358-1370.

Weir BS. The second National Research Council report on forensic DNA evidence. Am J Hum Genet. 1996; 59:497-500.

Yaghi L, Poras I, Simoes RT, et al. Hypoxia inducible factor-1 mediates the expression of the immune checkpoint HLA-G in glioma cells through hypoxia response element located in exon 2. Oncotarget. 2016; 7(39):63690-63707.

Zhao Y, Ren JL, Wang MY, et al. Codon 104 variation of p53 gene provides adaptive apoptotic responses to extreme environments in mammals of the Tibet plateau. Proc Natl Acad Sci U S A. 2013; 11051:20639-20644.

Internet resources section

http://www.soda-is.com, time of access 09:50 date 31.05.2017

http://www.worldclim.org, time of access 09:50 date 31.05.2017

http://www.altitude.org/air_pressure.php, time of access 09:50 date 31.05.2017

http://www.diva-gis.org, time of access 22:00 date 26.02.2019

http://www.fluxus-engineering.com, time of access 09:45 date 03.11.2018

http://www.geospiza.com/Products/finchtv.shtml, time of access 11:30 date 25.10.2018

Table 1. ***HLA-G*** **3'UTR haplotype frequencies observed in the populations (in %), according to altitude classification, and AMOVA results (*Fst* and *Fct* values).**

Populations(2n)	UTR-1	UTR-2	UTR-3	UTR-4	UTR-5	UTR-8	UTR-9	UTR-13	UTR-18	UTR-21	UTR-30
Highlands											
Amantani (56)	14.2 (8)	78.6 (44)	3.6 (2)	–	1.8 (1)	1.8 (1)	–	–	–	–	–
Anapia (30)	26.7 (8)	53.3 (16)	13.3 (4)	–	6.7 (2)	–	–	–	–	–	–
Aymara (10)	60.0 (6)	20.0 (2)	20.0 (2)	–	–	–	–	–	–	–	–
Cabanaconde (30)	40.0 (12)	33.3 (10)	13.4 (4)	10.0 (3)	3.3 (1)	–	–	–	–	–	–
Chivay (34)	52.9 (18)	35.3 (12)	8.8 (3)	–	3.0 (1)	–	–	–	–	–	–
Quechua (30)	40.0 (12)	40.0 (12)	13.4 (4)	–	3.3 (1)	–	–	–	3.3 (1)	–	–
Taquile (84)	34.5 (29)	35.7 (30)	21.4 (18)	–	7.2 (6)	–	–	–	–	–	1.2 (1)
Uros (42)	26.3 (11)	57.1 (24)	7.1 (3)	–	7.1 (3)	–	–	–	–	–	2.4 (1)
Yanke (18)	33.3 (6)	33.3 (6)	16.7 (3)	–	11.1 (2)	–	–	–	5.6 (1)	–	–
Mean frequency[1]	33.1	46.3	13.0	0.9	5.1	0.29	–	–	0.62	–	0.6
F_{ST}	0.016	**0.078**	-0.000	NA	-0.034	NA	NA	NA	0.069	NA	-0.029
(*p* value)	(0.235)	**(0.007)**	(0.436)		(0.952)				(0.087)		(0.918)

Lowlands											
Andoas (122)	22.1 (27)	45.9 (56)	15.6 (19)	0.8 (1)	15.6 (19)	–	–	–	–	–	–
Cinta Larga (14)	42.9 (6)	21.4 (3)	14.3 (2)	–	7.1 (1)	–	–	14.3 (2)	–	–	–
Guarani Kaiowá (12)	16.7 (2)	25.0 (3)	–	–	50.0 (6)	–	8.3 (1)	–	–	–	–
Guarani Ñandeva (12)	33.3 (4)	33.3 (4)	25.0 (3)	–	8.4 (1)	–	–	–	–	–	–
Lengua (22)	22.7 (5)	22.7 (5)	27.3 (6)	–	27.3 (6)	–	–	–	–	–	–
Xavante (40)	47.5 (19)	2.5 (1)	12.5 (5)	–	32.5 (13)	–	2.5 (1)	–	–	2.5 (1)	–
Xicrin (12)	75.0 (9)	16.7 (2)	–	–	8.3 (1)	–	–	–	–	–	–
Zoró (6)	–	–	33.3 (2)	–	66.7 (4)	–	–	–	–	–	–
Mean frequency[1]	29.7	30.6	15.6	0.00	21.6	–	0.8	0.8	–	0.4	–
F_{ST} (p value)	**0.077** **(0.050)**	**0.113** **(0.013)**	-0.028 (0.695)	NA	0.073 (0.078)	NA	-0.025 (0.472)	NA	NA	-0.025 (0.495)	NA
F_{CT} (p value)	-0.011 (0.646)	0.026 (0.165)	0.001 (0.399)	-0.014 (0.965)	**0.110** **(<0.001)**	0.004 (0.627)	0.005 (0.202)	-0.021 (0.451)	-0.017 (0.997)	0.005 (0.215)	0.011 (0.263)

[1]Weighted average. Number of haplotypes are displayed in brackets, and statistically significant values in **bold**. NA = not applicable.

Table 2. Correlation of *HLA-G* 3'UTR haplotype frequencies with environmental variables.

	UTR-1		**UTR-2**		**UTR-3**		**UTR-4**		**UTR-5**		**UTR-8**		**UTR-9**		**UTR-13**		**UTR-18**		**UTR-21**		**UTR-30**	
	Rho	*P*	Rho	*P*	Rho	*P*	Rho	*P*	Rho	*P*	Rho	*P*	Rho	*P*	Rho	*P*	Rho	*P*	Rho	*P*	Rho	
Latitude	0.02	0.95	-0.02	0.94	0.3	0.23	-0.04	0.87	-0.13	0.62	-0.06	0.8	-0.25	0.31	0.05	0.85	-0.05	0.86	-0.02	0.93	-0.09	0.72
Longitude	0.16	0.52	-0.45	0.06	-0.32	0.2	-0.13	0.61	0.44	0.07	-0.06	0.81	0.35	0.16	0.15	0.54	-0.2	0.43	0.32	0.2	-0.09	0.72
Altitude	0.15	0.56	**0.56**	**0.02**	-0.19	0.45	0.15	0.54	**-0.6**	**0.01**	0.23	0.37	-0.3	0.23	-0.25	0.32	0.36	0.14	-0.25	0.32	0.31	0.21
UV Irradiance	0.16	0.52	0.46	0.06	-0.19	0.45	0.2	0.44	**-0.52**	**0.03**	0.24	0.34	-0.25	0.32	-0.23	0.35	0.29	0.25	-0.1	0.69	0.32	0.19
O_2 mmHg	-0.11	0.66	**-0.57**	**0.01**	0.17	0.51	-0.16	0.52	**0.58**	**0.01**	-0.23	0.36	0.29	0.25	0.29	0.25	-0.36	0.15	0.24	0.35	-0.31	0.21
Annual Mean Temp.	-0.15	0.56	**-0.53**	**0.02**	0.14	0.57	-0.05	0.86	**0.59**	**0.01**	-0.22	0.39	0.27	0.27	0.23	0.35	-0.36	0.15	0.27	0.27	-0.29	0.24
Mean Diurnal Range	0.4	0.1	-0.12	0.63	-0.03	0.92	0.31	0.21	-0.29	0.25	-0.21	0.42	-0.16	0.54	-0.07	0.78	0.42	0.08	0.02	0.93	-0.13	0.61
Isothermality	0.2	0.43	**-0.56**	**0.02**	0.2	0.43	-0.2	0.43	0.27	0.28	-0.19	0.45	-0.1	0.69	0.29	0.24	0.03	0.92	0.31	0.21	-0.26	0.29
Temp. Annual Range	0.18	0.48	-0.18	0.48	0.21	0.41	0.11	0.65	-0.12	0.65	-0.22	0.39	0	0.99	-0.05	0.84	0.26	0.3	0.01	0.96	-0.17	0.5
Annual Precip	-0.23	0.36	-0.29	0.25	0.16	0.52	-0.3	0.23	0.43	0.07	-0.09	0.72	0.16	0.52	0.27	0.28	-0.32	0.19	0.12	0.64	-0.17	0.5

Statistically significant results are displayed in bold. Temp = temperature; Precip = precipitation; MmHg = millimeter of mercury. Mean diurnal range is the day-to-night temperature oscillation, whereas temperature annual range the summer-to-winter oscillation. Isothermality is a quantification of how large the mean diurnal range temperature oscillation is in comparison to the temperature annual range oscillation (a value of 1 represents a place here the diurnal temperature range is equal to the annual temperature).*P= p* value.

Table 3. Ewens-Watterson test results.

	Sample size	No. of haplotypes	Observed F value	Expected F value	Watterson F *p* value	Slatkin's exact *p* value
Highlands						
Amantani	29	5	0.43639	0.39353	0.94160	0.95810
Anapia	15	4	0.33333	0.41032	0.48000	0.48980
Aymara	5	3	0.44000	0.40572	1.0	1.0
Cabanaconde	15	4	0.65333	0.41246	0.13950	0.11070
Chivay	17	4	0.31488	0.42520	0.56550	0.69080
Quechua	14	3	0.35714	0.52552	0.13550	0.15500
Taquile	42	5	0.47619	0.42387	0.12390	0.23680
Uros	21	5	0.53288	0.36255	0.74820	0.86350
Yanque	10	5	0.26000	0.28266	0.28570	0.28570
Mean	18.66	4.2	0.42268	0.40465	*0.49110*	*0.53227*
s.d.	10.38	0.78	0.11502	0.05988	0.32673	0.33415
Lowlands						
Andoas	61	4	0.63666	0.53662	**0.03660**	**0.01800**
Cinta Larga	6	4	0.27778	0.30426	1.0	1.0
G. Kaiowa	6	3	0.33333	0.42964	0.60130	0.60130
G. Ñandeva	7	4	0.38776	0.32371	0.14050	0.14050
Lengua	10	4	0.34000	0.36587	**0.03760**	**0.03760**

Xavante	22	6	0.44628	0.30253	0.56910	0.70010
Xicrin	6	2	0.50000	0.63505	1.0	1.0
Zoro	3	2	0.55556	0.55556	1.0	1.0
Mean	15.12	3.62	0.43467	0.43165	*0.54814*	*0.56219*
s.d.	18.16	1.21	0.11503	0.12074	0.40378	0.40999

Statistically significant results are displayed in bold. The Ewens-Watterson test showed deviations compatible with the action of balancing selection in these populations (p<0,05). The mean of the highlands and lowlands populations are shown in italics.

Table S1. Variation sites at the HLA-G 3'UTR analyzed in this study.

Nucleotide change[a]	Accession number	Ancestral allele[c]	Minor allele[d]	Neanderthal[e]	Denisova[f]
14-bp indel	rs371194629	Ins	Ins	Del	Del
+3001 C/T	rs567747015	C	T	C	C
+3003 T/C	rs1707	T	C	T	T
+3010 C/G	rs1710	C	G	G	G/C
+3027 C/A	rs17179101	C	A	C	C/A
+3032 G/C	rs146339774	G	C	G	G
+3035 C/T	rs17179108	T	T	C	C/T
+3052 C/T	rs569057854	C	T	C	C
+3092 G/T	rs180827037	G	T	G	G
+3107 C/G	NA[b]	C	G	C	C
+3111 G/A	rs554784083	G	A	G/A	G
+3121 T/C	rs138249160	T	C	T	T
+3142 G/C	rs1063320	G	C	C	G/C
+3177 G/T	rs554076817	G	**T**	G	G
+3183 A/G	rs187320344	G	A	G	G
+3187 A/G	rs9380142	A	G	A	G/A
+3196 C/G	rs1610696	C	G	C	C
+3227 G/A	rs1233331	G	A	G	G

+3289 C/T	rs541542414	C	T	C	C

[a]Nucleotide positions of variation sites at 3'UTR are relative to the first ATG codon at exon 1, where the adenine is assigned position +1. [b]'NA' is used if the variation is not found in dbSNP and has no RefSeq identifier (rsid). [c]Ancestral allele according to Ensembl database. [d]Less common allele occurring in human populations. [e]Obtained from sequence data from six Neanderthal individuals, as displayed at the UCSC Genome browser (Briggs *et al*., 2007; Green *et al*., 2010). [f]Data from high-coverage sequencing of one Denisova individual from Altai mountains, Russia (Meyer *et al*., 2012; Reich *et al*., 2010).

Table S2. HLA-G haplotype composition regarding the variants considered in this study. Nomenclature according to Alvarez *et al*., 2009, Castelli *et al*., 2010, and Sabbagh *et al*., 2014.

	14 bp Indel	+3001 C/T	+3003 T/ C	+3010 C/G	+3027 A/C	3032 G/C	+3035 C/T	+3052 C/T	+3092 G/T	+3107 C/G	+3111 G/A	+3121 T/C	+3142 G/C	+3177 G/T	3183 A/G	+3187 A/G	+3196 C/G	+3227 G/A/C/T	+3289 A/G/C/T
	rs371194629	rs567747015	rs1707	rs1710	rs1717910	rs146339774	rs17179108	rs569057854	rs180827037	-	rs554784083	rs138249160	rs1063320	rs554076817	rs187320344	rs9380142	rs16610696	rs116248022 rs1233331	rs541542414
UTR-1	**Del**	C	T	**G**	C	G	**C**	C	G	C	G	T	**C**	G	G	**G**	C	G	C
UTR-2	Ins	C	T	C	C	G	**C**	C	G	C	G	T	G	G	G	A	**G**	G	C
UTR-3	**Del**	C	T	C	C	G	**C**	C	G	C	G	T	G	G	G	A	C	G	C
UTR-4	**Del**	C	**C**	**G**	C	G	**C**	C	G	C	G	T	**C**	G	G	A	C	G	C
UTR-5*	Ins	C	T	C	C	G	T	C	G	C	G	T	G	G	G	A	C	G	C
UTR-8	Ins	C	T	**G**	C	G	**C**	C	G	C	G	T	G	G	G	A	**G**	G	C
UTR-9	Ins	C	T	**G**	C	G	T	C	G	C	G	T	G	G	G	A	C	G	C
UTR-13	**Del**	C	T	C	C	G	T	C	G	C	G	T	G	G	G	A	C	G	C
UTR-18	**Del**	C	T	**G**	C	G	**C**	C	G	C	G	T	**C**	G	G	A	C	**A**	C
UTR-21	**Del**	C	T	**G**	C	G	**C**	C	G	C	G	T	G	G	G	A	C	G	C
UTR-30	Ins	C	T	**G**	C	G	**C**	C	G	C	G	T	**C**	G	G	**G**	C	G	C

Haplotype sequences are formed by the succession of polymorphisms from +2960 (14bp InsDel), +3001, +3003, +3010, +3027, +3032, +3035, +3052, +3092, +3107, +3111, +3121, +3142, +3177, +3183 +3187, +3196, +3227 and + 3289 along the 3'UTR region of the HLA-G gene (in the direction 5'→ 3').* Ancestral haplotype. +3027 rs1717910 A and +3121 rs138249160 C alleles occur just in Cabanaconde and Yanke, respectively (see Table S3). Haplotypes with these variants were excluded due to the low level of confidence of the haplotype inferred by the PHASE software.

Table S3. Native American allele frequencies (in %) from all polymorphic sites observed in the *HLA-G* 3'UTR region according to altitude classification. Number of haplotypes are displayed in brackets.

Populations (2n)	Ins/Del rs371194629 14bp*		rs1707 +3003		rs1710 +3010		rs1717910 +3027		17179108 +3035		rs138249160 +3121		rs1063320 +3142		rs9380142 +3187		rs1610696 +3196	
	Ins	Del	T	C	C	G	C	A	C	T	T	C	C	G	A	G	C	G
Highlands																		
Amantani (56)	0.82	0.18	1.00	-	0.84	0.16	1.00	-	0.98	0.02	1.00	-	0.14	0.86	0.86	0.14	0.20	0.80
Anapia (30)	0.60	0.40	1.00	-	0.73	0.27	1.00	-	0.93	0.07	1.00	-	0.27	0.73	0.73	0.27	0.47	0.53
Aymara (10)	0.25	0.75	1.00	-	0.33	0.67	1.00	-	1.00	-	1.00	-	0.60	0.40	0.40	0.60	0.80	0.20
Cabanaconde (30)	0.38	0.62	0.91	0.09	0.50	0.50	0.97	0.03	0.94	0.06	1.00	-	0.50	0.50	0.59	0.41	0.69	0.31
Chivay (34)	0.38	0.62	1.00	-	0.47	0.53	1.00	-	0.97	0.03	1.00	-	0.53	0.47	0.47	0.53	0.65	0.35
Quechua (30)	0.43	0.57	1.00	-	0.57	0.43	1.00	-	0.97	0.03	1.00	-	0.43	0.57	0.60	0.40	0.60	0.40
Taquile (84)	0.44	0.56	1.00	-	0.64	0.36	1.00	-	0.93	0.07	1.00	-	0.35	0.65	0.65	0.35	0.64	0.36
Uros (42)	0.67	0.33	1.00	-	0.71	0.29	1.00	-	0.93	0.07	1.00	-	0.29	0.71	0.71	0.29	0.43	0.57
Yanke (18)	0.45	0.55	1.00	-	0.65	0.35	1.00	-	0.85	0.15	0.95	0.05	0.35	0.65	0.70	0.30	0.70	0.30
Lowlands																		
Andoas (122)	0.61	0.39	0.99	0.01	0.77	0.23	1.00	-	0.84	0.16	1.00	-	0.23	0.77	0.78	0.22	0.54	0.46
Cinta Larga (14)	0.29	0.71	1.00	-	0.57	0.43	1.00	-	0.79	0.21	1.00	-	0.43	0.57	0.50	0.50	0.75	0.25
Guarani Kaiowá (12)	0.83	0.17	1.00	-	0.75	0.25	1.00	-	0.42	0.58	1.00	-	0.17	0.83	0.83	0.17	0.75	0.25
Guarani Ñandeva (12)	0.42	0.58	1.00	-	0.67	0.33	1.00	-	0.92	0.08	1.00	-	0.33	0.67	0.67	0.33	0.67	0.33

Lengua (22)	0.50	0.50	1.00	-	0.77	0.23	1.00	-	0.73	0.27	1.00	-	0.23	0.77	0.77	0.23	0.77	0.23
Xavante (40)	0.38	0.62	1.00	-	0.47	0.53	1.00	-	0.65	0.35	1.00	-	0.45	0.55	0.55	0.45	0.98	0.02
Xicrin (12)	0.08	0.92	1.00	-	0.25	0.75	1.00	-	0.92	0.08	1.00	-	0.75	0.25	0.25	0.75	1.00	-
Zoró (6)	0.62	0.38	0.88	0.12	1.00	0	1.00	-	0.50	0.50	1.00	-	-	1.00	1.00	-	0.88	0.12

* rs371194629 14 bases pairs (bp) Insertion (Ins) =GATTTGTTCATGCCT and/or Deletion (Del).

Table S4. *HLA-G* 3'UTR haplotype frequencies in worldwide populations.

			Groups											
Groups	**Populations**	**Total 2N**	**UTR-1**	**UTR-2**	**UTR-3**	**UTR-4**	**UTR-5**	**UTR-8**	**UTR-9**	**UTR-13**	**UTR-18**	**UTR-21**	**UTR-30**	**Others haplotypes**
Africa	GUI	120	19.17	38.33	6.67		1.67					3.33		30.83
	SER	478	11.51	34.31	27.20	5.44	9.41						0.21	11.92
	TOR	60	25.00	26.67	20.00	8.33	8.33							11.67
	YRI	176	13.07	27.84	28.98	13.07	7.95				1.14			7.95
	YAN	350	14.57	18.00	28.00	10.57	7.71	0.86		0.57		0.86	0.29	18.57
	LWK	192	24.48	32.29	11.98	8.85	8.33				3.13			10.94
	ASW	122	22.13	31.15	16.39	6.56	13.11				3.28			7.38
Europe	POR*	120	25.00	34.17	9.17		4.17					3.33		24.17
	IBS	28	32.14	39.29	7.14	10.71					7.14			3.57
	TSI	196	29.08	31.63	12.76	14.29	3.57				1.53			7.14
	CEU	170	39.41	25.88	8.82	15.29	3.53				1.18			5.88
	GBR	178	33.71	29.78	5.62	11.80	3.37				5.62			10.11
	FIN	186	36.02	21.51	6.45	28.49	1.61				1.08			4.84
Asia	JPT	178	24.72	19.66	49.44	0.56	1.69							3.93
	CHB	194	28.87	18.56	27.84	4.64	2.58							17.53
	CHS	200	43.00	6.50	26.50	2.00								22.00
America	SEB	310	25.81	24.19	12.58	13.23	9.35	1.29	0.65	0.00				12.90
	NEB	254	29.53	22.83	15.75	6.69	8.66			0.39				16.14

	CLM	120	23.33	24.17	23.33	14.17	7.50	5.00	2.50
	MXL	128	28.91	28.13	15.63	7.81	10.16	3.13	6.25
	PUR	110	29.09	13.64	19.09	12.73	11.82	7.27	6.36
Global		3870							

Abbreviations: 2N. total number of alleles; GUI. natives of Guinea-Bissau; SER. Serer from Niakhar. Senegal; TOR. Tori from Tori-Bossito. Benin; YRI. Yoruba from Ibadan. Nigeria; YAN. Yansi from Bandundu. Democratic Republic of the Congo; LWK. Luhya from Webuye. Kenya; ASW. people of African ancestry from the southwestern United States; POR. Portuguese; IBS. Iberian populations from Spain; TSI. Toscani from Italy; CEU. Utah residents with Northern and Westhern European ancestry; GBR. British from England and Scotland; FIN. Finnish from Finland; JPT. Japanese from Tokyo. Japan; CHB. Han Chinese from Beijing; CHS. Han Chinese from South China; SEB. Southeastern Brazilians from RibeirãoPreto. São Paulo. Brazil; NEB. Northeastern Brazilians from Recife. Pernambuco. Brazil; CLM. Colombians from Medellín. Colombia; MXL; people of Mexican ancestry from Los Angeles. California; PUR. Puerto Ricans from Puerto Rico. Table adapted from Sabbagh *et al.*. 2014.

Table S5. Bioclimatic variables evaluated in population of this study:

Population	Latitude	Longitude	Altitude	UV Irradiance	BIO1	BIO2	BIO3	BIO4	BIO5
Aymara	-19.25	-69.08	4312	253.4	32	181	71	254	194
Quechua	-12.33	-75.83	4433	242.9	32	142	785	181	854

Yanke	-15.64	-71.65	3820	251.0	93	182	74	243	430
Chivay	-15.63	-71.6	3820	251.0	83	183	73	248	453
Uros	-15.73	-69.92	3818	251.0	85	131	71	183	743
Anapia	-16.31	-68.85	3828	253.4	86	137	65	210	728
Amantani	-15.65	-69.71	3827	251.8	83	114	70	162	964
Taquile	-15.76	-69.68	3825	251.8	84	115	71	160	1134
Cabanaconde	-15.62	-71.98	3441	251.0	132	171	79	216	289
Andoas	-2.9	-76.4	235	1793	257	99	86	114	2552
Guarani Ñadeva	-23.8	-54.5	319	201.3	218	117	57	202	1593
Guarani Kaiowa	-23.1	-55.2	456	204.1	221	118	61	193	1566
Xavante	-14	-52.5	357	216.9	246	138	708	195	1534
Xikrin	-5.92	-51,00	300	201.4	249	114	829	138	1908
Cinta Larga	-11.17	-60,00	331	203.3	233	128	690	186	1945
Lengua	-22.75	-58.08	83	206.7	245	119	573	208	1184
Zoró	-10.33	-60.33	196	201.8	243	126	682	184	1924

Bio 1: Annual Mean Temperature, Bio 2: Mean Diurnal Range, Bio 3: Isothermality, Bio 4: Temperature Annual Range, Bio 5: Annual Precipitation. All temperature indexes are in °C × 10 and precipitation indexes are in mm. Irradiance is in W/m².

Figure 1. Geographic location of the Native American populations.

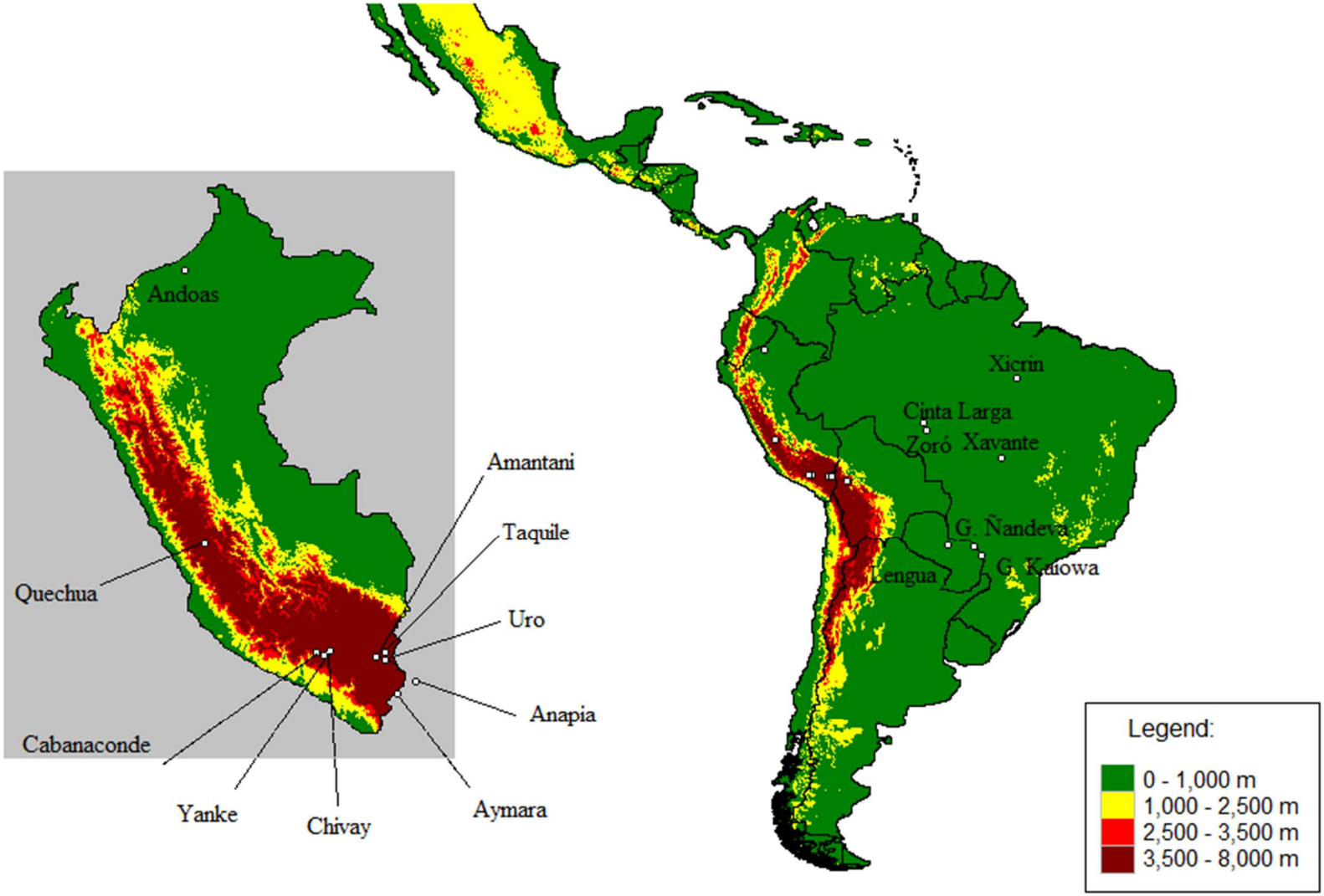
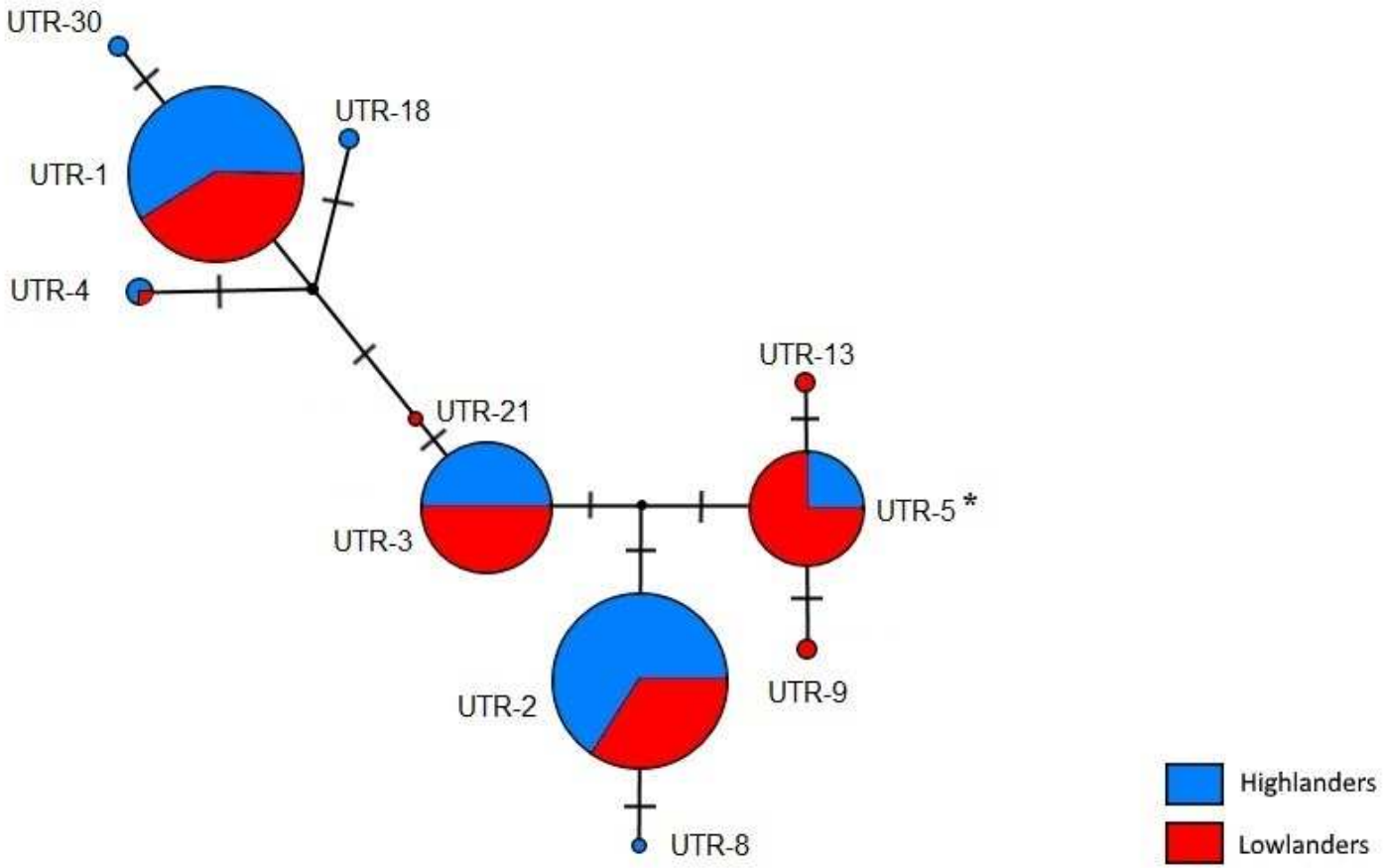


Figure 2. Median-joining haplotype network.



*UTR-5: Ancestral haplotype.

CAPÍTULO 6
DISCUSSÃO GERAL, REFERÊNCIAS E ANEXOS

6.1 Discussão geral

O estudo da adaptação para vida em altas altitudes é um tema bastante abordado na literatura científica mundial. Inserindo os termos "*high altitude AND human adaptation*" no PubMed encontramos 2.403 artigos científicos, sendo que o primeiro deles encontra-se publicado no ano de 1946, e já aborda o tema da hipóxia como um importante fator estressor do ambiente de altitude (Vannotti *et al*., 1946). Se restringirmos a busca às adaptações genéticas "*high altitude AND human genetic adaptation*" encontramos ainda 244 publicações, divididas entre as três populações de altitude do mundo, e alguns poucos estudos funcionais.

É importante salientar que mesmo com mais de duas centenas de publicações relativas à adaptação genética às altas altitudes, há consenso de que muito ainda precisa ser descoberto sobre genes e rotas envolvidas nesse fenômeno. Além disso, é imprescindível a necessidade de identificar as consequências funcionais das variantes genéticas identificadas em estudos com genes candidatos ou com varredura genômica (GWAS). Em recente revisão, Moore (Moore, 2017) destaca a necessidade de maior interdisciplinaridade para que haja um efetivo avanço no que tange a metodologia dos estudos de GWAS. Além disso, estudos com o objetivo de desvendar os mecanismos bioquímicos, fisiológicos e moleculares por trás de eventuais variantes genéticas funcionais precisam ser desenvolvidos (Bigham *et al*.,2014; Lorenzo *et al*.,2014; Song *et al*.,2014).

Nesta tese, são apresentados, inicialmente, dados a partir de uma abordagem de genes candidatos (Jacovas *et al*.,2015). Avaliando a diversidade dos cinco polimorfismos de nucleotídeo único (SNPs) localizados nos genes da via *TP53* (*TP53*, rs1042522; *MDM2*, rs2279744; *MDM4*, rs1563828; *USP7*, rs1529916; e *LIF*, rs929271) em 282 indivíduos Nativos Americanos oriundos de terras altas do Peru (Altiplano Andino, ≥ 2.500m) e terras baixas da América do Sul e Mesoamérica (<2.500m) identificamos os alelos ancestrais *USP7*-G, *LIF*-T e *MDM2*-T com evidências de que foram evolutivamente cooptados para o sucesso da colonização humana dos Andes.

Nossos resultados sugerem que uma regulação adequada da via *TP53* esteja relacionada à adaptação de altas altitudes nas populações andinas, mostrando que uma interação alélica específica permite não apenas a sobrevivência, mas principalmente, o sucesso reprodutivo dessas populações em ambientes hostis (Jacovas *et al*.,2015).

Partindo para uma abordagem exploratória mais ampla em um conjunto populacional mais extenso e heterogêneo, no artigo apresentado no capítulo 2 desta tese, realizamos a busca por regiões gênicas com sinal de seleção em populações do Altiplano Andino. Através de análises de dados de genomas (um total de 213.987 SNPs) de 322 Nativos Americanos oriundos de populações de terras altas (Altiplano Andino) e de terras baixas (Mesoamérica e Amazônia) nós identificamos cinco genes candidatos *TMEM38B* (*Transmembrane Protein 8B*), *AS3MT* (*Arsenite Methyltransferase*), *SP100* (*SP100 Nuclear Antigen*), *DUOX2* (*DualOxidase* 2) e *CLC* (*Charcot-Leyden Crystal Galectin*) que apresentavam sinal de seleção positiva em nativos dos Andes, Aymara e Quechua, sendo os três últimos (*SP100*, *DUOX2* e *CLC*) descritos pela primeira vez no nosso trabalho (Jacovas *et al*., 2018). Estes genes estão envolvidos na rede de *TP53*, e estão relacionados a rotas fisiológicas importantes para a resposta à hipóxia em grandes altitudes, como aquelas ligadas às adaptações do músculo esquelético e cardíaco, aumento da angiogênese e funções imunes na interface feto-materna.

Por exemplo, a proteína Sp100, codificada pelo gene *SP100,* está envolvida na condensação da cromatina e regulação transcricional com efeito estimulador na expressão gênica dependente de p53. Nós identificamos três SNPs com forte sinal de seleção natural neste gene, baseado na análise de PBS (*Population Branch Specific*). PBS basicamente estima, a partir das distâncias genéticas (medidas por F_{ST}) entre os três grupos populacionais considerados, se existe alelo(s) com frequência(s) extrema(s) no grupo andino em relação aos outros dois (Yi *et al*.,2010). Além disso, de acordo com nossas predições de expressão protéica *in silico*, um desses SNPs, rs13411586, causa uma expressão diferencial da proteína em humanos. Indivíduos homozigotos para o alelo selecionado (C) teriam sua produção Sp100 aumentada (Jacovas *et al*., 2018). Pode-se facilmente observar que a diferença na frequência média desse alelo é surpreendentemente grande entre dos grupos populacionais: andinos (frequência do alelo C=40%), mesoamericanos (4.5%) e amazônicos (5%), os dois últimos caracterizados por habitarem terras baixas.

Interessantemente, o gene *SP100* é diferencialmente expresso nos músculos esqueléticos. Estudos revelaram que um membro da via do HIF, HIF-1α, além de responder ao insulto promovido pela hipóxia, induzindo vários mecanismos de tolerância, como estabilização, angiogênese e autofagia (Zonneveld et al., 2019), desempenha um

papel importante na regulação da homeostase do oxigênio, incluindo tanto adaptações fisiológicas do músculo esquelético quanto do músculo cardíaco, em situações de redução de oxigênio devido ao esforço muscular. A exposição a grandes altitudes leva à redução da massa e do desempenho muscular (por exemplo, menor capacidade de trabalho e fadiga permanente), exceto quando se está evolutivamente adaptado a ela (Jacovas *et al*., 2018).

HIF-1α protege a sobrevivência celular durante o baixo suprimento de oxigênio, enquanto p53 promove a morte celular sob hipóxia. A razão para esses papéis aparentemente antagônicos pode estar na diferença da quantidade de oxigênio disponível; em uma condição normal, os níveis de p53 e HIF-1α são baixos, mas em hipóxia moderada, o nível de p53 permanece baixo, enquanto o nível de HIF-1α aumenta, protegendo as células ainda relativamente saudáveis da destruição. Em hipóxia severa, o acúmulo de p53 promove a repressão ou degradação de proteínas anti-apoptóticas como o HIF-1α, induzindo a apoptose das células sob estresse (Obacz *et al*.,2013 e Zhou *et al*.,2015). A proteína Sp100 é conhecida como moduladora da atividade da p53 e, sob hipóxia tecidual, isquemia, é regulada negativamente, levando à instabilidade genômica. Os andinos apresentam uma frequência relativamente alta do alelo C (rs13411586) que, em homozigose, aumentam a produção de Sp100 de acordo com o nosso teste de predição. Nosso resultado sugere então que houve uma solução evolutiva para manter Sp100 em um nível adequado em um ambiente com constante baixo nível de oxigênio (Jacovas *et al*., 2018).

Além disso, é possível especular que existe um equilíbrio intrincado no nível de expressão dos genes *SP100*, *TP53* e *HIF*-1α. No entanto, os exatos mecanismos celulares e moleculares pelos quais se dão essa adaptação, e como ocorre esse balanço ainda precisa ser investigado. Levando em conta seu relevante papel na via de hipóxia, os genes *SP100, TP53* e *HIF*-1α aparecem como ótimos alvos para estudo dos mecanismos moleculares e funcionais de resposta à hipóxia em populações nativas americanas que habitam há séculos as altas altitudes encontradas no Altiplano Andino.

Já o *DUOX2,* expresso em células epiteliais de vários tecidos participa da via do peróxido de hidrogênio (H_2O_2), que é necessária nas etapas finais da produção de hormônios tireoidianos. Ela também está envolvida formação das espécies reativas de oxigênio (ROS), um subproduto do metabolismo do oxigênio, mesmo sob condições fisiológicas normais. No entanto, diferentes condições estressoras podem aumentar a

produção de ROS, por exemplo, exposição a altas altitudes (hipóxia e altos índices de radiação ultravioleta) além de condições patológicas como o câncer. Salmeen e colaboradores forneceram evidências de que o DUOX2 desempenha um papel em um mecanismo de checkpoint dependente de p53 para a entrada no ciclo celular (Salmeen *et al*.,2010; Gupta *et al.*, 2012).

Adicionalmente, foi demonstrada que as espécies reativas de oxigênio desempenham um importante papel na angiogênese, o processo de formação de novos vasos sanguíneos. A angiogênese é um dos eventos principais na resposta à hipóxia (tolerância) e, portanto, pode ter um importante papel na adaptação à altitude dos nativos de terras altas, a fim de compensar uma plausível alteração na microcirculação (Ge *et al*.,2011; Buroker *et al*., 2012; Kim e Byzova*et al*., 2015).

O *CLC* codifica uma galectina que é reconhecida como uma lisofosfolipase e expressa em eosinófilos e basófilos (Ackerman *et al*., 2002). O único estudo funcional sobre essa proteína mostrou que em estado de hipóxia há o acúmulo de eosinófilos e de CLC, concomitante com um atraso na apoptose constitutiva, antagonizando o efeito pró-apoptótico normal de agentes que normalmente induzem a apoptose de eosinófilos (Porter *et al*., 2017). Ainda, a expressão alterada de genes da família das galectinas, incluindo CLC, foi implicada na emergência e progressão do câncer, destacando o papel das galectinas na proliferação celular através de programas de morte celular (Than *et al*., 2004; Su *et al*., 2018). Estudo avaliando parálogos da galectina em placenta ovina, mostraram que galectina-1 preveniu processos inflamatórios que prejudicam o feto, enquanto galectina-13, que tem a maior homologia à CLC, é um membro do grupo das chamadas "proteínas relacionadas à gravidez", devido às suas funções imunes especiais na interface feto-materna. Essas funções celulares fundamentais, já descritas para humanos e outros mamíferos placentários, podem indicar o caminho que conecta nossos achados de CLC e a pressão de seleção no clima hostil andino (Iglesias *et al*., 1998).

No entanto, mais estudos são necessários para confirmar o papel desses genes e variantes na adaptação às altas altitudes e o impacto funcional dessas na rede de *TP53*. Com este objetivo, ainda no ano de 2018, propomos um projeto intitulado “Caracterização dos Mecanismos Moleculares de Variante Envolvida na Adaptação à Vida em Altitude nos Andes”. Neste projeto, em parceria com a Profa. Úrsula Matte, propomos a investigação e o impacto da variante encontrada com maior sinal de seleção (Jacovas *et al*.,2018) -

rs13411586 no gene *SP100* - e desvendar as implicações funcionais em níveis transcricionais e traducionais na expressão de Sp100, p53 e HIF-1α *in vitro*, em condições de hipóxia, a fim de elucidar e compreender os seus mecanismos moleculares subjacentes. No momento, estratégias para o desenvolvimento dos experimentos estão sendo estabelecidas.

Outra molécula que nos despertou interesse de estudo no contexto de adaptações genéticas à vida em altas altitudes foi o HLA-G. Conforme apresentado inicialmente no Capítulo 1. tópico 1.5. desta tese, o HLA-G desempenha, reconhecidamente, um papel de suma importância na gestação humana (Nagamatsu *et al*.,2004). Além disso, já foi descrito níveis plasmáticos alterados de HLA-G em alpinistas do Monte Everest, sugerindo um papel desta molécula na adaptação a altas altitudes (Bourguignon *et al*.,2010), pelo menos num contexto de curto-tempo e com reversibilidade. Varios autores tem também indicado seu papel no câncer (Garziera et al., 2017). Em comum, esses microambientes seja normal (gestação) ou patológico (câncer), bem como o aqui chamado "macro-ambiente" de altas altitudes, tem a hipóxia. Uma cascata de mecanismos de tolerância a hipoxia (angiogênese, por exemplo) tem sido descritos, dentros os quais vários desencadeados por HLA-G.

Assim, sequenciamos a região regulatória 3'UTR do *HLA-G* de 301 Nativos Americanos, divididos em populações localizadas em altas altitudes (>2.500m) e terras baixas (<2.500m). Um total de 11 haplótipos foram observados, sendo que nenhum haplótipo novo foi descrito. Destes, encontramos três (UTR-8. -18 e -30) exclusivos em indivíduos de terras altas e outros três (UTR-9. -13 e -21) exclusivamente encontrados em terras baixas. Com relação à estruturação populacional, interessantemente observamos que a variância observada entre os grupos (F_{CT}) é marcadamente maior (11%) para o UTR-5, um haplotipo ancestral e ligado a baixa produção de HLA-G, indicando uma notável e significativa diferença entre as distribuições haplotípicas de UTR-5 entre populações de terras altas e baixas. Análises adicionais nos mostraram correlação estatística entre UTR-2 com diversas variáveis climáticas, demonstrando que este haplótipo pode estar relacionado à proteção contra o ambiente hostil das altas altitudes. A presença alta de UTR-5 nas terras baixas precisa ser ainda melhor investigada. Vimos ainda que uma parte importante dos polimorfismos na região promotora do HLA-G estão também presentes em neandertais e no espécime de Denisova. Esse resultado indica que estes são polimorfismos ancestrais

(trans-SNPs), mantidos a milhares de anos, possivelmente por seleção balanceadora, não obstante não termos conseguido sinal desse fenômeno com o teste aqui empregado.

Tendo em vista os aspectos associados às adaptações genéticas para a vida nas altas altitudes do Altiplano Andino, nossos dados corroboram a ideia de que não há somente uma única resposta adaptativa ao estresse devido às altas altitudes. Apesar disto, sugerimos aqui que combinações alélicas que promovem uma regulação adequada da via *TP53* (vários genes), bem como de *HLAG-G*, desencadeiem cascatas de eventos e sinalizações relacionadas à adaptação (sobrevivência e reprodução) às altas altitudes, pelo menos nas populações andinas (Jacovas et al., 2015, 2018 e capítulo V da presente tese). Algumas dessas variantes, inclusive, são antigas, possivelmente presente em espécies de homininios ancestrais dos *Homo sapiens* e arcaicos (neandertal e Denisova), sendo mantidas por seleção baleneadora até os dias atuais, pelo menos nos humanos modernos.

Interessantemente em Jacovas et al. (2015), referimos relatórios do governo peruano que mostra que a taxa de complicações gestacionais e pós-parto no povo andino Aymara chega a ser 5 vezes menor do que a média nacional. No entanto, estes também teriam mais susceptibilidade a uma doença comum nos dias atuais, o câncer (Volskarides et al. 2018). Baseado nessas informações e nos achados da presente tese, uma hipótese pode ser sugerida: um repertório genético complexo, envolvendo uma delicada combinação de antigos e derivados alelos tem permitido a tolerância a hipóxia em nível macro ambiental (alta atitude) e microambiental normal (placenta e outras condições correlacionadas à gestação). Como resultado, os andinos estão bem adaptados a uma longa vida em altas altitudes. Por outro lado, esse mesmo repertório genético pode ser desfavorável a esses indivíduos nos dias atuais, considerando um contexto microambiental patológico específico, uma vez que fornece ferramentas para a célula cancerígena, normalmente sob hipóxia, sobreviver e se multiplicar. Vale ressaltar que, mesmo considerando um sistema de reparo supostamente mais eficiente nos indivíduos que viem em terras altas, necessário para evitar muitos danos ao DNA e as células devido à alta incidência de UV e temperaturas frias, este é relativamente ineficiente para controlar a incidência de câncer, também associada a muitas mutações tumorais somáticas, subsequentes à mutação de origem (Petljak et al., 2019). Isso faz sentido, considerando que a grande maioria dos casos de câncer humano surge após o período reprodutivo, tornando essa doença, como a conhecemos hoje, irrelevante em um contexto evolutivo.

No entanto, vale destacar que os exatos mecanismos sejam eles moleculares, celulares ou fisiológicos envolvidos nessa tipo de adaptação ainda merecem ser melhor investigados. Assim, estudos funcionais são chaves para corroborar ou descartar as hipóteses aqui sugeridas. Nossos próximos passos serão neste sentido.

6.2 Referências bibliográficas

Ackerman SJ, Liu L, Kwatia MA,Savage MP, Leonidas DD, Swaminathan GJ, Acharya KR (2002) Charcot-Leyden crystal protein (galectin-10) is not a dual function galectin with lysophospholipase activity but binds a lysophospholipase inhibitor in a novel structural fashion. J Biol Chem 277:14859-14868.

Ashur-Fabian O, Avivi A, Trakhtenbrot L, Adamsky K, Cohen M, Kajakaro G, Joel A, Amariglio N, Nevo E, Rechavi G (2004) Evolution of p53 in hypoxia-stressed Spalax mimics human tumor mutation. Proc Natl Acad Sci U S A 10133:12236-12241.

Bigham AW, Julian CG, Wilson MJ, Vargas E, Browne VA, Shriver MD, Moore LG (2014) Maternal PRKAA1 and EDNRA genotypes are associated with birth weight. and PRKAA1 with uterine artery diameter and metabolic homeostasis at high altitude. Physiol Genomics 46:687-697.

Bourguignon M, Yaghi L, Flajollet S, Radanne-Krawice I, Rouas-Freiss N, Lugrin D, Richalet JP, Carosella ED, Moreau P (2010) Increased soluble human leukocyte antigen-G levels in peripheral blood from climbers on Mount Everest. Hum Immunol 71: 1105-1108.

Buroker NE, Ning XH, Zhou ZN, Li K, Cen WJ, Wu XF, Zhu WZ, Scott CR, Chen SH (2012) AKT3, ANGPTL4, eNOS3, and VEGFA associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. Int J Hematol 96:200-213.

Dumont P, Leu JI, Della Pietra AC 3rd. George DL, Murphy M (2003) The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet 33:357-365.

Garziera M, Scarabel L, Toffoli G. Hypoxic Modulation of HLA-G Expression through the Metabolic Sensor HIF-1 in Human Cancer Cells (2017) J Immunol Res 4587520.

Ge RL, Mo VY, Januzzi JL, Jin G, Yang Y, Han S,Wood MJ, Levine BD (2011) B-type natriuretic peptide, vascular endothelial growth factor, endothelin-1, and nitric oxide synthase in chronic mountain sickness. Am J Physiol Heart Circ Physiol 300:H1427-1433.

Gupta SC, Hevia D, Patchva S, Park B, Koh W, Aggarwal BB (2012) Upsides and downsides of reactive oxygen species for cancer: the roles of reactive oxygen species in tumorigenesis, prevention. and therapy. Antioxid Redox Signal 16:1295-1322.

Iglesias MM, Rabinovich GA, Ivanovic V, Sotomayor C, Wolfenstein-Todel C (1998) Galectin-1 from ovine placenta, Amino-acid sequence, physicochemical properties and implications in T-cell death. Eur J Biochem 252: 400-407.

Jacovas VC, Couto-Silva CM, Nunes K, Lemes RB, de Oliveira MZ,Salzano FM, Bortolini MC, Hünemeier T (2018) Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. Sci Rep 8:12733.

Jacovas VC, Rovaris DL, Peréz O, de Azevedo S, Macedo GS, Sandoval JR, Salazar-Granara A, Villena M, Dugoujon JM, 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:e0137823

Kim MJ, Ryu JC, Kwon Y, Lee S, Bae YS, Yoon JH, Ryu JH (2014) Dual oxidase 2 in lung epithelia is essential for hyperoxia-induced acute lung injury in mice. Antioxid Redox Signal 21:1803-1818.

Kim Y and Byzova TV (2015) Oxidative stress in angiogenesis and vascular disease. Blood 123:625–632.

Lorenzo FR, Huff C, Myllymäki M, Olenchock B, Swierczek S, Tashi T, Gordeuk V. Wuren, Ri-Li G, McClain DA *et al*., (2014) A genetic mechanism for Tibetan high-altitude adaptation. Nat Genet 46:951–956.

Moore LG (2017) Human Genetic Adaptation to High Altitudes: Current Status and Future Prospects. Quat Int 461:4-13.

Nagamatsu T, Fujii T, Yamashita T, Miki A, Kanai T, Kusumi M, Osuga Y, Kozuma S, Taketani Y (2004) Hypoxia does not reduce HLA-G expression on extravillous cytotrophoblasts. J Reprod Immunol 63:85–95.

Obacz J, Pastorekova S, Vojtesek B and Hrstka R (2013) Cross-talk between HIF and p53 as mediators of molecular responses to physiological and genotoxic stresses. Mol Cancer 12:93.

Petljak M, Alexandrov LB, Brammeld JS, et al (2019) Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell. 176:1282-1294.e20.

Porter LM, Cowburn AS, Farahi N, Deighton J, Farrow SN, Fiddler CA, Juss JK, Condliffe AM, Chilvers ER (2017) Hypoxia causes IL-8 secretion. Charcot Leyden crystal formation, and suppression of corticosteroid-induced apoptosis in human eosinophils. Clin Exp Allergy 47:770-784.

Salmeen A, Park BO and Meyer T (2010) The NADPH oxidases NOX4 and DUOX2 regulate cell cycle entry via a p53-dependent pathway. Oncogene 29:4473-4484.

Song D, Li LS, Arsenault PR, Tan Q, Bigham AW, Heaton-Johnson KJ, Master SR, Lee FS (2014) Defective Tibetan PHD2 binding to p23 links high altitude adaption to altered oxygen sensing. The Journal of biological chemistry 289:14656-14665.

Su J, Wang Y, Si Y, Gao J, Song C, Cui L, Wu R, Tai G, Zhou Y (2018) Galectin-13, a different prototype galectin. does not bind β-galacto-sides and forms dimers via intermolecular disulfide bridges between Cys-136 and Cys-138. Sci Rep 8:980.

Than NG, Pick E, Bellyei S, Szigeti A, Burger O, Berente Z, Janaky T, Boronkai A, Kliman H, Meiri H *et al*., (2004) Functional analyses of placental protein 13/galectin-13. Eur J Biochem 271:1065-1078.

Vannotti A (1946) The adaptation of the cell to effort. altitude and to pathological oxygen deficiency. Schweiz Med Wochenschr 76:899-903.

Voskarides K (2018) Combination of 247 Genome-Wide Association Studies Reveals High Cancer Risk as a Result of Evolutionary Adaptation. Mol Biol Evol 35:473-485.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS *et al*., (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75-78.

Zhao Y, Ren JL, Wang MY, Zhang ST, Liu Y, Li M, Cao YB, Zu HY, Chen XC, Wu CI., *et al*., 2013. Codon 104 variation of p53 gene provides adaptive apoptotic responses to extreme environments in mammals of the Tibet plateau. Proc Natl Acad Sci U S A 11051:20639-20644.

Zhou CH, Zhang XP, Liu F and Wang W (2015) Modeling the interplay between the HIF-1 and p53 pathways in hypoxia. Sci Rep 5:13834.

Internet resources section

http://www.dge.gob.pe/publicaciones/pub_asis/asis26.pdf, p. 165;

http://www.dge.gob.pe/portal/docs/intsan/asis2012.pdf, p. 76;

6.3 ANEXOS

6.4 Outras publicações científicas

6.4.1 *Novel genetic associations and gene-gene interactions of chemokine receptor and chemokine genetic polymorphisms in HIV/AIDS*

Publicado como: Valverde-Villegas JM, de Medeiros RM, de Andrade KP,**Jacovas VC**, Dos Santos BR, Simon D, de Matos Almeida SE,Chies JAB. Novel genetic associations and gene-gene interactions of chemokine receptor and chemokine genetic polymorphisms in HIV/AIDS. AIDS. 2017 Jun 1;31(9):1235-1243. doi: 10.1097/QAD.0000000000001491. PubMed PMID: 28358741.

Abstract:

Objective: To investigate the influence of candidate polymorphisms on chemokine receptor/ligand genes on HIV infection and AIDS progression (HIV/AIDS). Design: Fifteen polymorphisms of the CCR3, CCR4, CCR5, CCR6, CCR8, CXCR3, CXCR6, CCL20, CCL22 and CXCL10 genes were analysed in 206 HIV-positive patients classified as rapid progressors (n=40), or nonrapid progressors (n=166), and in 294 HIV-seronegative patients. Methods: The polymorphisms were genotyped using minisequencing. Genetic models were tested using binomial logistic regression; nonparametric multifactor dimensionality reduction (MDR) was used to detect gene-gene interactions. Results: The CCR3 rs3091250 [TT. adjusted odds ratio (AOR): 2.147, 95% confidence interval (CI) 1.076-4.287, P=0.030], CCR8 rs2853699 (GC/CC. AOR: 1.577. 95% CI 1.049-2.371. P=0.029). CXCL10 rs56061981 (CT/TT. AOR: 1.819. 95% CI 1.074-3.081. P=0.026) and CCL22 rs4359426 (CA/AA. AOR: 1.887. 95% CI 1.021-3.487. P=0.043) polymorphisms were associated with susceptibility to HIV infection. The CCL20 rs13034664 (CC. OR: 0.214. 95% CI 0.063-0.730. P=0.014) and CCL22 rs4359426 (CA/AA. OR: 2.685. 95% CI 1.128-6.392. P=0.026) variants were associated with rapid progression to AIDS. In MDR analyses revealed that the CXCL10 rs56061981 and CCL22 rs4359426 combination was the best model. with 57% accuracy (P=0.008) for predicting susceptibility to HIV infection. Conclusion: Our results provide new insights into the influence of candidate chemokine receptor/ligand polymorphisms and significant evidence for gene-gene interactions on HIV/AIDS susceptibility.

6.4.2 *A tale of agriculturalists and hunter-gatherers: Exploring the thrifty genotype hypothesis in native South Americans*

Publicado como: Reales G, Rovaris DL,**Jacovas VC**, Hünemeier T, Sandoval JR, Salazar-Granara A, Demarchi DA, Tarazona-Santos E, Felkl AB, Serafini MA, Salzano FM, Bisso-Machado R, Comas D, Paixão-Côrtes VR,Bortolini MC. A tale of agriculturalists and hunter-gatherers: Exploring the thrifty genotype hypothesis in native South Americans. Am J Phys Anthropol. 2017 Jul;163(3):591-601. doi: 10.1002/ajpa.23233. Epub 2017 May 2. PubMed PMID: 28464262.

Abstract:

Objectives: To determine genetic differences between agriculturalist and hunter-gatherer southern Native American populations for selected metabolism-related markers and to test whether Neel's thrifty genotype hypothesis (TGH) could explain the genetic patterns observed in these populations. Materials and Methods: 375 Native South American individuals from 17 populations were genotyped using six markers (APOE rs429358 and rs7412; APOA2 rs5082; CD36 rs3211883; TCF7L2 rs11196205; and IGF2BP2 rs11705701). Additionally, APOE genotypes from 39 individuals were obtained from the literature, AMOVA, main effects, and gene-gene interaction tests were performed. Results: We observed differences in allele distribution patterns between agriculturalists and hunter-gatherers for some markers. For instance, between-groups component of genetic variance (FCT) for APOE rs429358 showed strong differences in allelic distributions between hunter-gatherers and agriculturalists ($p = 0.00196$). Gene-gene interaction analysis indicated that the APOE E4/CD36 TT and APOE E4/IGF2BP2 A carrier combinations occur at a higher frequency in hunter-gatherers. but this combination is not replicated in archaic (Neanderthal and Denisovan) and ancient (Anzick, Saqqaq, Ust-Ishim, Mal'ta) hunter-gatherer individuals. Discussion: A complex scenario explains the observed frequencies of the tested markers in hunter-gatherers. Different factors, such as pleotropic alleles. rainforest selective pressures, and population dynamics, may be collectively shaping the observed genetic patterns. We conclude that although TGH seems a plausible hypothesis to explain part of the data, other factors may be important in our tested populations.

6.4.3 *Melanoma Research: Skin Pigmentation Polymorphisms Associated with Increased Risk of Melanoma in Individuals from Southern Brazil*

Artigo submetido à revista científica *Melanoma Research*. Autores: Larissa Brussa Reis, Renato Marchiori Bakos, Fernanda Sales Luiz Vianna, Gabriel de Souza Macedo, **Vanessa Cristina Jacovas,** André Santos, Sidney Santos, Lúcio Bakos, Patricia Ashton- Prolla.

Abstract:

Melanoma is the most aggressive type of skin cancer, associated with several environmental and genetic risk factors, and it originates in melanocytes, the pigment-producing skin cells. Single nucleotide polymorphisms (SNPs) in pigmentation genes have been described in melanoma risk modulation but our knowledge in the field is still limited. This investigation was based on the effect of SNPs in four pigmentation genes – TYR (rs1126809), HERC2 (rs1129038), SLC24A5 (rs1426654), and SLC45A2 (rs16891982) on risk for melanoma using a multivariate logistic regression and a Multifactor Dimensionality Reduction (MDR) analysis. in a case-control approach with individuals from Southern Brazil. with and without population substructure. Our results show three of the four SNPs studied was associated, two in dominant model: HERC2 rs1129038AA or a SCL24A5 rs1426654AA increased the risk of melanoma [OR= 2.094 (95% CI: 1.106 - 3.966). P= 2.3x10-2], and [OR= 7.126 (95% CI: 1.873 – 27.110). P= 4x10-3], respectively, and the SLC45A2 rs16891982 in an additive model with protection for melanoma twice higher when two allele C is inherited [OR= 0.081 (95% CI: 0.008 – 0.782). P= 3x10-2]. In addition, the MDR analyze found that the combination of rs1426654 AA and rs16891982 GG genotypes is associated with higher risk for melanoma (P=3x10-3). These results contribute to the current knowledge indicating that SNPs in pigmentation genes can reflect the risk of melanoma.