UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS
INSTITUTO DE BIOCIÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR – PPGBM

Relação genótipo → fenótipo e a pigmentação humana: aspectos evolutivos e sua aplicação na genética forense

CAIO CESAR SILVA DE CERQUEIRA

Tese de doutorado submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM) da Universidade Federal do Rio Grande do Sul (UFRGS), como requisito parcial para obtenção do grau de Doutor em Ciências.

Orientadora: Profa. Dra. Maria Cátira Bortolini
Co-orientadora: Dra. Tábita Hünemeier

Porto Alegre – RS, Agosto de 2013.

Este trabalho foi desenvolvido no Laboratório de Evolução Humana e Molecular (LEHM) do Departamento de Genética/Instituto de Biociências da Universidade Federal do Rio Grande do Sul (UFRGS).

INSTITUIÇÕES FINANCIADORAS

CNPq
CAPES
FAPERGS
FUNDAÇÃO *Levehulme* (Reino Unido)

Dedico esta tese aos meus orientadores/mentores durante a Graduação em Ciências Biológicas, Mestrado em Genética e Biologia Molecular e Doutorado em Genética e Biologia Molecular, bem como às pessoas que me auxiliaram na trajetória acadêmico-científica, formando o profissional que sou atualmente.

"Negra, vermelha, branca, amarela...

Cor da pele é banalidade.

A caveira que somos é atestado de igualdade".

(Anita Costa Prado, 2005 *apud* Salzano, 2012 – Genômica e Evolução)

AGRADECIMENTOS

À Professora Maria Cátira Bortolini, por sua capacidade formidável de orientação, por sua humildade e inteligência, e pelos momentos agradáveis vividos durante todo o doutorado.

À Co-orientadora Tábita Hünemeier, pelo companheirismo, amizade, discussões acaloradas, risadas proporcionadas e uma inigualável intensidade de relacionamento experimentado ao longo desse tempo.

À Francis Zambra, pelo apoio incondicional, pela dedicação altruística concedida, pelo carinho extraordinário oferecido de forma crescente ao longo do tempo, pela sinceridade e, principalmente, pela cumplicidade e amor vividos intensamente.

Aos companheiros que tiveram a paciência de dividir um lar comigo nos últimos anos Ana Cristoff, Felipe Oliveira e Francis Zambra.

Aos amigos e colegas da sala 129, tanto os atuais quanto os que já passaram por lá, são eles: Ágatha Xavier, Carla Volasko, Carlos Alencar, Clênio Machado, Carlos Eduardo Amorim, Eli Vieira, Franciele Santos, Isadora Lima, Luisa Lemos, Luiza Saldanha, Luana Maciel, Lucas Viscardi, Michael Kent, Pâmela Paré, Pedro Vargas, Rafael Bisso, Rodrigo Dornelles, Tábita Hünemeier, Vanessa Rodrigues e Virgínia Ramallo, pelas brincadeiras e momentos agradáveis vividos num ambiente profissional bastante prolífico.

Aos pesquisadores companheiros do Projeto Candela, especialmente aos Professores Dr. Andres Ruiz-Linares e Dr. Rolando González-José, os quais eu tive um imenso prazer de conhecer pessoalmente.

Aos outros amigos e colegas que ajudaram direta ou indiretamente durante as coletas das amostras utilizadas neste trabalho e que ainda não foram citados acima: Ana Braga, Ana Angélica, Sélia Heck, Danaê Longo, Dário Bezerra, Ivonélia Alcântara, Lavínia Schüller-Faccini, Maria Odete Quinto, Rosilene Paim e Uilson Souza.

Ao Coordenador administrativo do PPGBM e amigo Elmo Cardoso, por sua simplicidade, sua esplêndida capacidade de trabalho e as conversas trocadas durante os momentos de café na secretaria do PPGBM.

Aos amigos Diógenes Marques (Guga), Otávio Carvalho (Chetelba) e Otávio Gaya (Tavinho), pelas risadas e momentos divertidíssimos compartilhados.

Por fim e não menos importante, aos familiares mais próximos, que sempre me apoiaram e torceram por mim, inclusive àqueles que não estão mais presentes no convívio, mas presentes na memória: Zenita Silva (Avó Materna – *in memoriam*), Onaldo Silva (Avô

Materno – *in memoriam*), Andiara Silva (Mãe), Edson Reis (Padrasto – *in memoriam*), Cauan Antônio (Irmão), Permínio Augusto (Tio), Cristina Santos (Tia), Pedro Ernesto (Tio), Ivonilda Lopes (Tia), Zenaldo Pereira (Tio), Patrícia Silva (Prima-Irmã), Neilton Cerqueira (Pai), Normélia Fernandes (Madrasta), Neila Cerqueira (Irmã) e Ney Cerqueira (Irmão).

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS

ACTH (*Adrenocorticotropic Hormone*) – Hormônio Adrenocorticotrópico

ASIP (*Agouti Signaling Protein*) - Proteína de Sinalização Agouti

cAMP (*Cyclic Adenosine Monophosphate*) – Monofosfato Cíclico de Adenosina

CANDELA - Consórcio para Análise da Diversidade e Evolução Latino-Americana

Cenpat – Centro Nacional Patagônico

CIELab (*Commission International d'Eclairage Lab*) – Comissão Internacional de Iluminação Lab

CIEXYZ (*Commission International d'Eclairage XYZ*) – Comissão Internacional de Iluminação XYZ

Conicet (*Consejo Nacional de Investigaciones Científicas y Técnicas*) – Conselho Nacional de Investigações Científicas e Técnicas

CREB (*cAMP Responsive Element Binding Protein*) – Proteína de ligação ao elemento responsivo do cAMP

DCT (*Dopachrome Tautomerase*) - Dopacromo Tautomerase

DNA (*Deoxyribonucleic Acid*) – Ácido Desoxirribonucléico

DSM II (*Derma Spectrometer II*) – Espectrofotômetro Dérmico II

EVCs (*Externally visible characteristics*) - Características Visíveis Externamente

EXOC2 (Exocyst Complex Component 2) – Componente Complexo *Exocist* 2

FDP (*Forensic DNA phenotyping*) - Fenotipagem Forense por DNA

GWAS (*Genome Wide Association Studies*) – Estudo de Associação por Varredura Genômica

GWS (*Genome Wide Scan*) – Varredura Genômica

HERC2 (*HECT and RLD domain containing E3 ubiquitin protein ligase 2*) – Domínio HECT e RLD contendo a proteína ligase 2 ubiquitina E3

HSV/HSB (*Hue - Saturation - Value - Brightness*) – Matiz-Saturação-Valor-Brilho (sistema de cor)

IE - Índice de Eritema

IL-6 (*Interleukin 6*) – Interleucina 6

IM - Índice de Melanina

INDEL – Polimorfismo de inserção/deleção

IPE (*Iris Pigment Epithelium*) - Epitélio de Pigmento da Íris

IRF4 (*Interferon Regulatory Factor 4*) – Fator regulatório interferon 4

KITLG (*KIT Ligand*) – Ligante KIT

LD (*Linkage Disequilibrium*) - Desequilíbrio de Ligação

LEDs (*Ligh-Emitting Diode*) - Diodos Emissores de Luz

LEF-1 (*Lymphoid Enhancer-binding Factor 1*) – Fator de ligação/intensificador do linfoide 1

MATP (*Membrane-associated Transporter Protein*) – Proteína Transportadora Associada à Membrana

MC1R (*Melanocortin 1 Receptor*) - Receptor 1 de Melanocortina

MITF (*Microphthalmia Transcription Factor*) - Fator de Transcrição Microftalmia

OCA2 (*Oculocutaneous Albinism II*) – Albinismo Óculo-Cutâneo II

PAX3 (*Paired Box 3*) – *Paired Box 3*

POMC (*Proopiomelanocortin*) - Pro-opiomelanocortina

RGB (*Red-Green-Blue*) – Vermelhor-Verde-Azul (sistema de cor)

ROS (*Reactive Oxygen Species*) - Espécies Reativas de Oxigênio

SLC24A4 (*Solute Carrier Family 24, Member 4*) – Família Transportadora de Soluto 24, Membro 4

SLC45A2 (*Solute Carrier Family 45, Member 2*) - Família Transportadora de Soluto 45, Membro 2

SNPs (*Single nucleotide polymorphisms*) - Polimorfismos de Nucleotídeo Único

SOX1 (*SRY - Sex Determining Region Y - Box 1*) - Região do Y Determinante do Sexo (SRY) – *Box* 1

SOX9 (*SRY - Sex Determining Region Y - Box 9*) – Região do Y Determinante do Sexo (SRY) – *Box* 9

STRs (*Short Tandem Repeats*) – Repetições curtas consecutivas

TCF (*Transcription Factor*) – Fator de Transcrição

TCLE - Termo de Consentimento Livre e Esclarecido

TNF-α (*Tumour Necrosis Factor Alpha-like*) – Fator de Necrose Tumoral Alfa

TYR (*Tyrosinase*) - Tirosinase

TYRP1 (*Tyrosinase-related Protein 1*) - Proteína Relacionada à Tirosinase 1

UVB (*Ultraviolet Radiation* B) – Radiação Ultravioleta B

UVR (*Ultraviolet Radiation*) - Radiação Ultravioleta

α-MSH (*Alpha-melanocyte-stimulating Hormone*) – Hormônio Estimulante de Melanócitos Alfa

RESUMO

As populações brasileiras são caracterizadas por seus distintos (e extensos) padrões de mestiçagem. Justamente por apresentarem diferenças na proporção da herança genética africana, ameríndia e européia, é esperado que as conexões entre genótipos e fenótipos complexos como é o caso da cor de pele, olhos e cabelos, por exemplo, não sejam idênticas entre diferentes populações brasileiras, ou entre estas e aquelas encontradas em seus grupos ancestrais. Isto é corroborado pela consistente evidência de que grupos geográficos humanos diferem quanto ao *background* genético relacionado à pigmentação, refletindo-se, portanto, nas suas populações derivadas. Elucidar este *background* em amostras miscigenadas como a nossa é de fundamental importância para as áreas médica, evolutiva e forense. Na presente tese buscamos entender o que há de mais recente na literatura científica relacionada com a pigmentação humana, bem como produzir conhecimento de extrema relevância neste contexto em populações brasileiras com diferentes histórias demográficas. Para isto, estudamos o efeito de 18 SNPs com pigmentação da pele, olhos e cabelos na amostra de 563 voluntários brasileiros do consórcio internacional CANDELA (Consórcio para análise da diversidade e evolução latino-americana). Além disso, tentamos estabelecer um método de predição de fenótipos de pigmentação com a possível aplicação em genética forense e evolução humana. Na análise de associação dos 18 SNPs com fenótipos de pigmentação, observamos 3 (três) marcadores genéticos (*HERC2*rs1129038, *SLC24A5*rs1426654 e *SLC45A2*rs16891982) consistentemente associados com cor da pele, olhos e cabelos, nas distintas populações brasileiras investigadas. Este é o primeiro passo para compreender melhor o contexto biológico da produção de melanina e entender as relações genótipo-fenótipo da pigmentação em populações miscigenadas. Com a compreensão dos marcadores genéticos relacionados com pigmentação e também com o aperfeiçoamento do método de predição de fenótipos de pigmentação desenvolvido no presente trabalho, é possível dar o primeiro passo para a criação de uma metodologia de predição de fenótipos de cor da pele, olhos e cabelos em populações com histórias demográficas miscigenadas como a nossa, com possível aplicação nas áreas forense, médica e evolutiva.

ABSTRACT

Brazilian populations are characterized by their distinct (and extensive) admixture patterns. Precisely because there are differences in the proportion of African, Amerindian and European genetic heritage, it is expected that the connections between genotypes and complex phenotypes such as skin, eyes and hair color, are not identical among different Brazilian populations, or between these and those found in their ancestral groups. This is corroborated by consistent evidence that human geographic groups differ in genetic background related to pigmentation, reflecting thus in the respectives populations derived. Elucidating this background in admixed samples such as ours is of fundamental importance to the medical, forensic and evolutionary fields. In this thesis we seek to understand what is latest in the scientific literature related to human pigmentation, as well as producing knowledge extremely relevant in this context in Brazilian populations with different demographic histories. For this, we studied the effect of 18 SNPs with pigmentation of the skin, eyes and hair in a sample of 563 Brazilian volunteers of the international consortium called CANDELA (Consortium for the analysis of the diversity and evolution of Latin America). Also, we try to establish a method for predicting pigmentation phenotypes with possible application in forensic genetics and human evolution. In association analysis of 18 SNPs with pigmentation phenotypes, we observed 3 (three) genetic markers (*HERC2*rs1129038, *SLC24A5*rs1426654 and *SLC45A2*rs16891982) consistently associated with skin, eye and hair color, in distinct Brazilian populations investigated. This is the first step to better understand the biological context of the production of melanin and know the connexions between genotype-phenotype of the pigmentation in admixed populations. With the understanding of the genetic markers related to pigmentation as well as with the improvement of the method of predicting phenotypes developed in the present work, it is possible to take the first step toward establishing a methodology for predicting phenotypes of skin, eye and hair color in populations with demographic histories like ours, with possible application in forensic, medical and evolutionary fields.

1. INTRODUÇÃO

1.1 Considerações Gerais

A análise da abordagem populacional com genomas completos ainda é incipiente (*The 1000 Genomes Project Consortium*, 2010), diferente da já consolidada identificação e análises envolvendo somente sítios variáveis, sendo os mais estudados os polimorfismos de nucleotídeo único (*SNPs*, do inglês, *single nucleotide polymorphisms*). O projeto HapMap (*http://hapmap.ncbi.nlm.nih.gov/*), pioneiro neste tipo de abordagem é um exemplo, e a contribuição na geração de dados populacionais, com acesso rápido e fácil através da internet, tem também trazido importantes subsídios para várias áreas do conhecimento biológico. Contextualizado nesse universo da enorme quantidade de dados gerados rapidamente e com relativo baixo custo, surgem projetos com grandes metas e objetivos desafiadores. Por exemplo, estudar a relação de cada gene em particular com seus respectivos fenótipos é algo nada trivial devido à enorme complexidade das conexões entre genótipo → fenótipo, em especial quando os traços não apresentam herança mendeliana. Além disso, existe uma enorme variação normal dentro e entre populações, sendo que a extensão desta variação bem como as conexões entre cada variante com seu respectivo fenótipo ainda estão longe de ser conhecidas.

Diante desta problemática desafiadora, vários tipos de abordagens vêm sendo empregadas para o estudo de traços complexos, cada uma delas apresentando diferentes níveis de elucidação. A maioria destes estudos, no entanto, têm sido realizados em populações européias e seus descendentes, e alguns em asiáticos (Graf *et al.*, 2005, 2007; Norton *et al.*, 2007; Stokowski *et al.*, 2007; Edwards *et al.*, 2010), não abrangendo deste modo o grande espectro da variação normal humana. Além disso, é conhecido que associações encontradas em um grupo populacional não podem ser facilmente transpostas para outro, pois diferentes padrões de desequilíbrio de ligação (do inglês *LD*, *linkage disequilibrium*) podem ser determinantes em estudos de associação, daí a importância de estudos em populações específicas, com diferentes históricos evolutivos e demográficos.

As populações brasileiras são caracterizadas por seus distintos (e extensos) padrões de mestiçagem (Marrero *et al.*, 2005, 2007; Hünemeier *et al.*, 2007; Gonçalves *et al.*, 2008; Guerreiro-Junior *et al.*, 2009, para citar somente exemplos envolvendo nosso grupo de pesquisa). Justamente por apresentarem diferenças na proporção da herança genética africana, ameríndia e européia, é esperado que as conexões entre genótipos e fenótipos

complexos como é o caso da cor de pele, olhos e cabelos, por exemplo, não sejam idênticas entre diferentes populações brasileiras, ou entre estas e aquelas encontradas em seus grupos ancestrais.

1.2 Contextualizando o presente trabalho no Consórcio Internacional CANDELA

O trabalho de tese aqui desenvolvido faz parte de um projeto maior conduzido por nosso grupo de pesquisa intitulado "Fenótipo, ancestralidade genômica e dinâmica de mestiçagem no Brasil" (aprovado pelos comitês de ética da UFRGS, HCPA e UESB – Anexos I-III), estando este, por sua vez, vinculado ao Consórcio Internacional denominado CANDELA (*Consórcio para Análise da Diversidade e Evolução Latino-Americana; www.ucl.ac.uk/silva/candela*). O CANDELA merece destaque aqui por se tratar do mais abrangente estudo interdisciplinar com populações Latino-americanas envolvendo diversidade biológica, em nível de fenótipos visíveis (cor de pele, olhos e cabelos, medidas antropométricas, entre outros) e genótipos, juntamente com aspectos sócio-culturais envolvidos na dinâmica de mestiçagem e construção da identidade e pertencimento (os dados coletados neste projeto podem ser vistos nos Anexos IV-IX). Amostras e dados oriundos de cinco países latino-americanos (México, Colômbia, Peru, Chile e Brasil) foram coletados e estão sendo analisados através da ampla rede de projetos vinculados ao CANDELA. O coordenador geral do consórcio é o pesquisador Andres Ruiz-Linhares (*University College London*), enquanto as Profa. Maria Cátira Bortolini e Lavínia Schuler-Faccini são as responsáveis pelo mesmo no Brasil. O Prof. Rolando González-José (Cenpat/Conicet - Argentina), por sua vez, coordena a parte do consórcio relacionada com os dados antropométricos.

Parte da amostra brasileira coletada de acordo com o protocolo CANDELA foi investigada durante o desenvolvimento desta tese de doutorado. Detalhes a respeito deste protocolo serão descritos nos itens a seguir, sempre que pertinente, e/ou na sessão de Material e Métodos dos respectivos artigos científicos, bem como no Apêndice I desta tese.

1.3 Pigmentação humana

A cor dos olhos foi uma das primeiras características consideradas para ilustrar o conceito de herança mendeliana em humanos (Sturm & Frudakis, 2004). Desde o início dos estudos quantitativos de Genética Humana, a cor dos olhos tem servido como um modelo de traço facilmente observável que poderia se ajustar a uma regra mendeliana de herança dominante-recessiva (Sturm & Larsson, 2009). Foi no início do século XX que surgiram os dois primeiros relatos, um dos Estados Unidos e outro do Reino Unido, investigando a natureza da herança da cor dos olhos. Davenport & Davenport (1907) e Hurst (1908) estudaram a herança desta característica em famílias durante três gerações e ambos concluíram que a cor do olho castanho era um traço dominante, sendo o olho azul, uma característica recessiva. Infelizmente, como em muitos traços fenotípicos, este modelo não contempla a complexidade encontrada na natureza e, embora não seja comum, há relatos (tão antigos quanto os estudos iniciais de cor dos olhos) de que pais com olhos azuis podem também produzir crianças com olhos castanhos (Holmes & Loomis, 1909; Wright, 1918). O exemplo não-mendeliano de transmissão da cor da íris dos pais para os filhos, combinado com a natureza quantitativa da pigmentação da íris, já é suficiente para indicar que a herança deste traço não se ajusta a um modelo simplista.

Deste modo, assim como a cor da pele e cabelos, entre outros traços visíveis com claro fator hereditário, a cor da iris dos olhos deve ser mencionada como característica de herança multifatorial, condicionada por vários genes, além de outros fatores (Sturm *et al*., 1998; Badano & Katsanis, 2002). Entretanto, a despeito deste conhecimento, é ainda comumente ensinado em escolas através de livros didáticos o modelo equivocado e simplista de que a cor do olho castanho é sempre dominante sobre o azul (Sturm & Frudakis, 2004). Os últimos autores sugerem inclusive que o uso da cor do olho como um paradigma para a ação gênica dominante e recessiva deve ser evitado no ensino de genética para todos os graus de escolaridade.

O exemplo anterior ilustra que a pigmentação humana como um todo pode ser considerada como um traço complexo que apresenta variações dentro e entre populações (Makova & Norton, 2005). Sabe-se também, desde a época de Charles Darwin, que a variação encontrada na pigmentação humana pode ser explicada pela ação da seleção natural

(Darwin, 1871; Loomis, 1967; Post *et al.*, 1975; Branda & Eaton, 1978; Przeworski *et al.*, 2000; Mackintosh, 2001; para uma revisão mais recente, consulte Jablonski & Chaplin, 2010, 2012; Quillen & Shriver, 2011; Jablonski, 2012).

O primeiro estudo que mostrou o mapa da distribuição da pigmentação da pele humana data de 1953, e foi feito pelo italiano Renato Biasutti. Apesar de ser antigo, o mapa de Biasutti ainda é útil para descrever o padrão geográfico geral da pigmentação da pele nas populações humanas, o qual apresenta uma forte correlação com a latitude (Biasutti, 1953). Basicamente, a cor da pele tende a ser mais escura nas áreas tropicais e equatoriais, do que em áreas localizadas mais distantes da linha do equador (Figura 1). Um estudo realizado posteriormente, utilizando uma estimativa mais precisa de pigmentação baseada em reflectometria, mostrou o mesmo padrão geográfico da pigmentação da pele descrito no estudo de 1953 (Relethford, 1997). Versões mais atuais do mapa de Biasutti podem ainda ser vistas em Jablonski & Chaplin (2000), Chaplin (2004) e Parra (2007). O fator subjacente que explica esta notável relação entre pigmentação da pele e latitude parece ser a intensidade da luz UVR (radiação ultravioleta), que é grande na linha do equador e diminui progressivamente com o aumento da latitude.

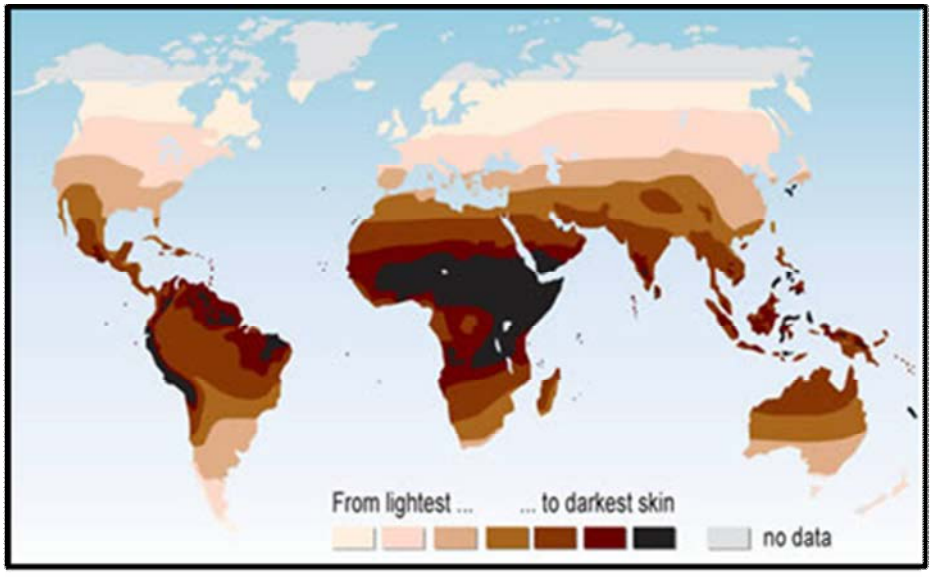


Figura 1. Média de cor da pele encontrada em nativos dos diferentes continentes. Adaptado de Chaplin (2004).

Jablonski & Chaplin (2000) exploraram com detalhes os fatores evolutivos e o padrão geográfico da pigmentação humana. Estes autores observaram uma forte correlação entre a reflectância da pele e os níveis de UVR incidentes nas regiões geográficas analisadas. Em

contraste com a pigmentação da pele, a variação da cor dos cabelos e olhos é muito mais restrita geograficamente. A maioria das populações humanas possui cabelos e olhos escuros. Cabelos ruivos/avermelhados e loiros são mais comumente encontrados em populações européias, embora esses traços sejam também observados em populações nativas da Austrália e Melanésia (Birdsell, 1993). A cor dos olhos azuis, verdes e mel são também predominantemente encontrados na Europa, com mais rara presença em nativos de outros continentes (Sturm & Frudakis, 2004). O estudo de Jablonski & Chaplin (2000) foi o primeiro a medir diretamente a intensidade da UVR em diferentes regiões a fim de testar as hipóteses relacionadas à evolução da pigmentação da pele humana. Independente do papel benéfico da síntese aumentada da vitamina D, os efeitos da radiação ultravioleta B (UVB) sobre a pele são universalmente prejudiciais. A supressão da sudorese e a subsequente interrupção da termorregulação devido ao dano induzido pela queimadura solar às glândulas de suor são os efeitos imediatos mais sérios da exposição excessiva à UVB (Daniels, 1964; Pandolf *et al*., 1992). Detalhes destes processos serão vistos nos parágrafos seguintes.

A melanina é um polímero natural que atua como um bloqueador solar natural e é especialmente efetiva em proteger contra os danos dos baixos comprimentos de onda da radiação eletromagnética (~300 nm), os quais são os maiores prejudiciais para o DNA e proteínas (Rees, 2003). Este biopolímero possui uma grande capacidade para absorver a luz visível, a UVR e a radiação ionizante, e também de neutralizar espécies reativas de oxigênio (ROS - *reactive oxygen species*), criadas quando estes agentes interagem com as células da pele (Ito, 2003; Rees, 2003). Sabe-se que a UVR danifica o DNA e os constituintes das membranas das células, causando uma cascata tóxica de eventos que produzem ROS, o que desestabiliza as reações químicas normais nas células (Cleaver & Crowley, 2002; Ballaré *et al*., 2011). Estes processos são bastante atenuados por um tipo de melanina conhecida como eumelanina (pigmento no espectro castanho/preto), especialmente quando esta se encontra presente próximo à superfície da pele (Nielsen *et al*., 2006). Em contraste, a feomelanina (pigmento no espectro vermelho/amarelo), mais comum na pele clara, parece aumentar o risco de estresse oxidativo nos melanócitos (Jablonski, 2012). Estas diferentes propriedades da eumelanina e da feomelanina com relação ao potencial de oxidação e fototoxicidade ajudam a explicar a alta incidência de câncer em indivíduos com alto conteúdo de feomelanina (Hill & Hill, 2000; Takeuchi *et al*., 2004; Samokhvalov *et al*., 2005; Ye *et al*., 2006).

O efeito inicial da exposição ao UVR é a queimadura solar, que é caracterizada por eritema (vermelhidão), edema, e possivelmente dor e formação de pequenas bolhas na pele (Rees, 2004). Queimaduras solares sucessivas podem causar danos nas glândulas de suor, podendo resultar no risco aumentado de infecção nas células da pele danificada. Desta maneira, em ambientes tropicais, local inicial em que a nossa espécie evoluiu, a pele escura com altas quantidades de eumelanina absorvendo UVR teria sido vantajosa, enquanto que a pele clara, mais propensa às queimaduras, sofria com danos nas glândulas de suor, e as infecções resultantes teriam um provável efeito adverso (Jablonski & Chaplin, 2000; Rees, 2004). Portanto, o conhecimento de que a exposição solar excessiva pode levar ao câncer de pele é sustentado pelo dano da UVR ao DNA, que interfere nos produtos gênicos que normalmente inibiriam o crescimento de células cancerosas (Halliday, 2005; Ullrich, 2005). Em pessoas com pele escura, há evidências de que a UVR promova a distribuição dos melanossomos tanto nas áreas supranucleares dos queratinócitos para protegê-los contra o dano ao DNA, como também na região da epiderme superior (mais externa) a fim de proteger a região mais profunda da epiderme, a qual possui células-tronco de melanócitos e queratinócitos (Tadokoro *et al.*, 2005). Isso pode explicar as diferenças no risco de câncer de pele entre as pessoas (Rees, 2004). Por esta razão, alguns autores têm indicado que a seleção natural favoreceu a pele escura em regiões equatoriais e tropicais para proteger a pele contra o câncer (Robins, 1991). Entretanto, os tipos mais sérios de câncer se desenvolvem após a idade reprodutiva; portanto, é improvável que o câncer de pele teve um papel significativo na evolução da cor da pele (Jablonski & Chaplin, 2000). Contudo, é importante mencionar neste aspecto, que albinos adolescentes - ou próximos da idade adulta - habitantes de áreas com alta incidência de UVR tipicamente desenvolvem lesões pré-malignas ou câncer de pele, sendo que, em alguns lugares como Nigéria e Tanzânia, menos de 10% dos albinos sobrevivem além dos 30 anos de idade (Robins, 1991; Yakubu & Mabogunje, 1993; Rees, 2003).

A luz solar não prejudica somente a pele humana em si, mas também a produção de alguns nutrientes essenciais, como o folato, que é extremamente sensível à UVR (Off *et al.*, 2005). Este nutriente é necessário para a síntese e reparo de DNA, e sua deficiência pode resultar em complicações durante a gravidez e múltiplas anormalidades fetais, incluindo defeitos de tubo neural, espinha bífida e anencefalia (Bower *et al.*, 1989; Lucock, 2000; Off *et al.*, 2005). Além disso, a fotólise do ácido fólico pode levar a erros meióticos durante a

produção do esperma (Mathur *et al.*, 1977). Isto sugere que em áreas geográficas com altos níveis de UVR, os indivíduos com pele clara estariam mais propensos à deficiência de folato do que àqueles com a pele escura. Devido à importância do folato em vários processos biológicos, é provável que a manutenção de níveis ótimos deste nutriente esteja sob a influência da seleção natural. Portanto, a baixa prevalência de deficiência grave do folato (Lawrence, 1983; Lamparelli *et al.*, 1988) e defeitos de fechamento de tubo neural (Carter, 1970; Elwood & Elwood, 1980; Wiswell *et al.*, 1990; Buccimazza *et al.*, 1994; Shaw *et al.*, 1994) observada entre nativos americanos e afro-americanos, mesmo entre indivíduos com deficiências nutricionais, pode ser atribuída a pele escura dos mesmos, a qual protegeria contra a fotólise do folato. Jablonski & Chaplin (2000), defendem que a proteção contra a fotólise de alguns nutrientes, em especial do folato, seria a razão principal para que a seleção natural favorecesse uma pele altamente pigmentada nas pessoas que viviam em ambientes com níveis elevados de radiação UVB.

Embora os efeitos da UVR sobre a pele são, em maior parte, prejudiciais, como vimos nos parágrafos anteriores, existe uma importante exceção: a radiação UVB é essencial para a síntese de vitamina D. Enquanto algumas fontes de alimento possuem quantidades substanciais de vitamina D (por exemplo, óleo de bacalhau), a sua síntese cutânea a partir do 7-deidrocolesterol via exposição solar é a principal fonte desta vitamina (Holick, 2003, 2005). Merece destaque que a vitamina D participa no metabolismo dos ossos e sua deficiência resulta em raquitismo em crianças, e osteomielite em adultos. Outras funções da vitamina D incluem imunorregulação, bem como regulação da diferenciação e proliferação celular, regulação da absorção de cálcio pelo intestino, bem como no funcionamento normal do pâncreas, cérebro e coração (Holick, 2004; Norman, 2008; Köstner *et al.*, 2009). Em condições não-adequadas de UVR disponível, sabe-se que a pele com grandes concentrações de melanina torna-se não adaptativa, pois não permite uma suficiente síntese de vitamina D (Loomis, 1967; Holick *et al.*, 1981; Clemens *et al.*, 1982; Cavalli-Sforza *et al.*, 1994). Além disso, se a duração da exposição à UVR não for suficiente para síntese de vitamina D, alguns indivíduos podem ter alto risco de deficiência desta vitamina e suas manifestações. Este fenômeno tem sido demonstrado em migrantes da Índia para centros urbanos no Reino Unido (Henderson *et al.*, 1987) e em migrantes da Etiópia para Israel (Fogelman *et al.*, 1995), indicando que os indivíduos com pele escura são altamente suscetíveis à deficiência de vitamina D causada pela mudança no regime de UVR. Os casos de toxicidade por

hipervitaminose são extremamente raros (Pandita *et al*., 2012), e sabe-se que a exposição solar, por si só, não resulta em toxicidade por vitamina D (Holick, 2007). Um dado interessante é que, para catalizar a síntese de uma quantidade equivalente de vitamina D, as pessoas que possuem índices de melanina de moderado a elevados requerem de 2 a 6 vezes maiores quantidades de radiação UV do que aqueles indivíduos que possuem pele mais clara (Holick *et al*., 1981; Clemens *et al*., 1982).

Para finalizar, a idéia de que a seleção natural poderia ser responsável pela variação da cor da pele tem uma história longa e uma grande quantidade de evidências. Deste modo pode-se sintetizar sugerindo que a distribuição da pigmentação observada nas populações nativas dos diferentes continentes é resultado da seleção natural que teria favorecido a pele escura para a proteção contra queimaduras solares e contra a destruição do folato em áreas onde a exposição da UVR é alta, e a seleção para pele mais clara em regiões distantes do equador para maximizar a síntese de vitamina D.

Com relação à evolução da cor dos cabelos e olhos, as informações na literatura científica são bem mais escassas. Sabe-se que alguns marcadores moleculares associados ao desenvolvimento da íris também estão associados com algumas doenças neurológicas como a Síndrome de Down (Donaldson, 1961), Neurofibromatose tipo 1 (Lee & Stepheson, 2007) e Síndrome de Gillespie (Ticho *et al.*, 2006; Marien *et al.*, 2008). Este fato indica que o desenvolvimento da íris e do cérebro podem estar ligados (Sturm & Larsson, 2009). Além disso, genes expressos na íris estariam também associados com alguns parâmetros psiquiátricos (Larsson *et al*., 2007). O conhecimento desta relação íris-cérebro foi um adicional incentivo para estudos sobre os aspectos evolutivos envolvidos com a variação da cor dos olhos em humanos. Merece destaque que o alto grau de esclera branca exposta na morfologia natural do olho humano, ausente em outros primatas, tem sido proposto como elemento importante na sinalização corporal envolvida nas interações pessoais (Kobayashi & Kohshima, 2001). Existem ainda sugestões indicando que as diferenças na pigmentação da íris poderiam influenciar na acuidade visual em ambientes de baixa luminosidade (Sturm & Larsson, 2009), no tempo de reação e na aptidão para esportes específicos (Rowe & Evans, 1994), bem como com timidez na infância (Coplan *et al*., 1998).

Para a cor dos cabelos, não encontramos na literatura científica um fator evolutivo que ajude a explicar a variabilidade em populações européias e outras populações ao redor do

mundo. Existem alguns estudos associando cabelos ruivos com doença de Parkinson e síndrome de Tourette (Gao *et al.*, 2009; Sterling-Levis & Williams, 2009), mas sem mencionar uma explicação evolutiva subjacente a esta característica. Como muitos marcadores genéticos de cor de cabelos são também associados com cor dos olhos e pele, é provável que a diversificação destes últimos traços ao redor do globo tenham também conduzido à diversificação da cor dos cabelos.

Na seminal obra-prima "*The descent of man*", de Charles Darwin (1871), aparece a sugestão de que as diferenças populacionais observadas para vários traços humanos, incluindo a pigmentação, poderia ser resultado da seleção sexual. Tal proposição tem ganhado reforço com trabalhos recentes, pois existiriam evidências que a pigmentação é um importante critério para a escolha de um companheiro (a), o que levou alguns autores a postularem que a seleção sexual teria sido um importante fator na manutenção da variação não somente da cor da pele, mas também da cor dos cabelos e olhos na nossa espécie (Diamond, 1991; Aoki, 2002; Frost, 2006). Outras hipóteses evolutivas para explicar a amplitude de variação da pigmentação humana incluem ainda resistência às infecções bacterianas, parasíticas e virais conferidas pela pele escura (Mackintosh, 2001) e resistência aos danos devido a ambientes gelados conferido pela pele clara (Post *et al*., 1975), bem como coevolução gene-cultura (Khan & Khan, 2010). Além disso, sabe-se que a melanina (em particular, a eumelanina) também exerce um efeito protetor retirando radicais livres reativos e outros oxidantes do organismo (Bustamante *et al.*, 1993; Krol & Lieber, 1998). Mais detalhes sobre estas hipóteses, bem como de outras podem ser vistas em Jablonski & Chaplin (2000; 2010; 2012), Parra (2007), Cerqueira *et al*. (2011; Apêndice II - capítulo de livro publicado por nosso grupo de pesquisa durante o desenvolvimento da presente tese), e em Jablonski (2012).

Independente dos mecanismos que originaram e mantêm a variação na pigmentação humana, merece destaque o fato da pigmentação, em especial a cor da pele, ser considerada a característica física mais visível das pessoas (Jablonski & Chaplin, 2012). Sendo assim, a cor da pele tem sido historicamente utilizada para definir grupos de identidade, e deste modo reforçar argumentos políticos e sócio-culturais, tanto de segregação quanto de inclusão, como a atual política de cotas em instituições públicas no Brasil (considerações sobre esta questão podem ser vistas no Apêndice III).

1.4 Mecanismos fisiológicos da pigmentação humana

Com relação aos aspectos fisiológicos, as diferenças na pigmentação entre os indivíduos são explicadas pelos seguintes fatores: a quantidade e o tipo de melanina sintetizada nos melanócitos (razão eumelanina/feomelanina), bem como a forma e distribuição dos melanossomos (Sturm *et al*., 1998; Rees, 2004; Parra, 2007). Os melanócitos são células derivadas da crista neural que migram para a epiderme durante o primeiro trimestre de gestação e produzem melanina. A melanina é uma mistura de polímeros com diferentes propriedades químicas, enzimaticamente derivada da tirosina por uma série complexa de passos intermediários (Ito, 2003), que veremos adiante. A melanina é sintetizada dentro de estruturas celulares conhecidas como melanossomos (estruturas similares aos lisossomos) e, quando madura, são passadas para queratinócitos adjacentes (Parra, 2007). A cor da pele visível é o resultado da melanina nos queratinócitos, mas que são sintetizadas nos melanócitos. Estudos genéticos em camundongos têm fornecido informações consideráveis sobre melanogênese, e um número grande de importantes produtos gênicos que afetam a alteração do balanço de vários tipos de polímeros – e, portanto, a natureza da melanina – tem sido identificados (Barsh, 1996; Jackson, 1997; Rees, 2003).

Em humanos, a regulação da melanogênese e a distribuição da melanina diferem na pele, cabelos e íris. Na pele, os melanócitos localizados na camada basal da epiderme transferem os melanossomos para queratinócitos adjacentes através de estruturas dendríticas, e os queratinócitos eventualmente migram para as camadas superiores da epiderme (Rees, 2003) (Figura 2a). Dentro dos queratinócitos, os melanossomos são tipicamente reunidos ao redor do núcleo celular, fornecendo proteção contra a UVR. A pele clara é enriquecida com a feomelanina, e os melanossomos tendem a ser menos pigmentados, menores em tamanho e empacotados em grupos. A pele escura, por sua vez, possui mais eumelanina, e os melanossomos são mais pigmentados, maiores, e distribuídos como unidades individuais (Alaluf *et al*., 2002).

Nos cabelos, o bulbo capilar é o único local de produção da melanina (Figura 2b). Neste local, melanócitos com grandes quantidades de melanina transferem os melanossomos para queratinócitos imaturos no entorno, os quais finalmente se diferenciarão e migrarão para formar o fio de cabelo pigmentado (Slominski *et al*., 2005). Nos cabelos, a melanogênese

somente ocorre durante o ciclo inicial de crescimento capilar (o período da formação da haste do cabelo que dura, em média, de 3 a 5 anos), enquanto na epiderme, o fenômeno parece ser contínuo (Ortonne & Prota, 1993). Já foi descrito que indivíduos com cabelos castanho-escuros/pretos possuem uma razão eumelanina/feomelanina elevada, os indivíduos com cabelos loiros/castanho-claros apresentam níveis de eumelanina e feomelanina similares, e os indivíduos com cabelos ruivos/avermelhados possuem os mais altos níveis de feomelanina (Ortonne & Prota, 1993; Rees, 2003).

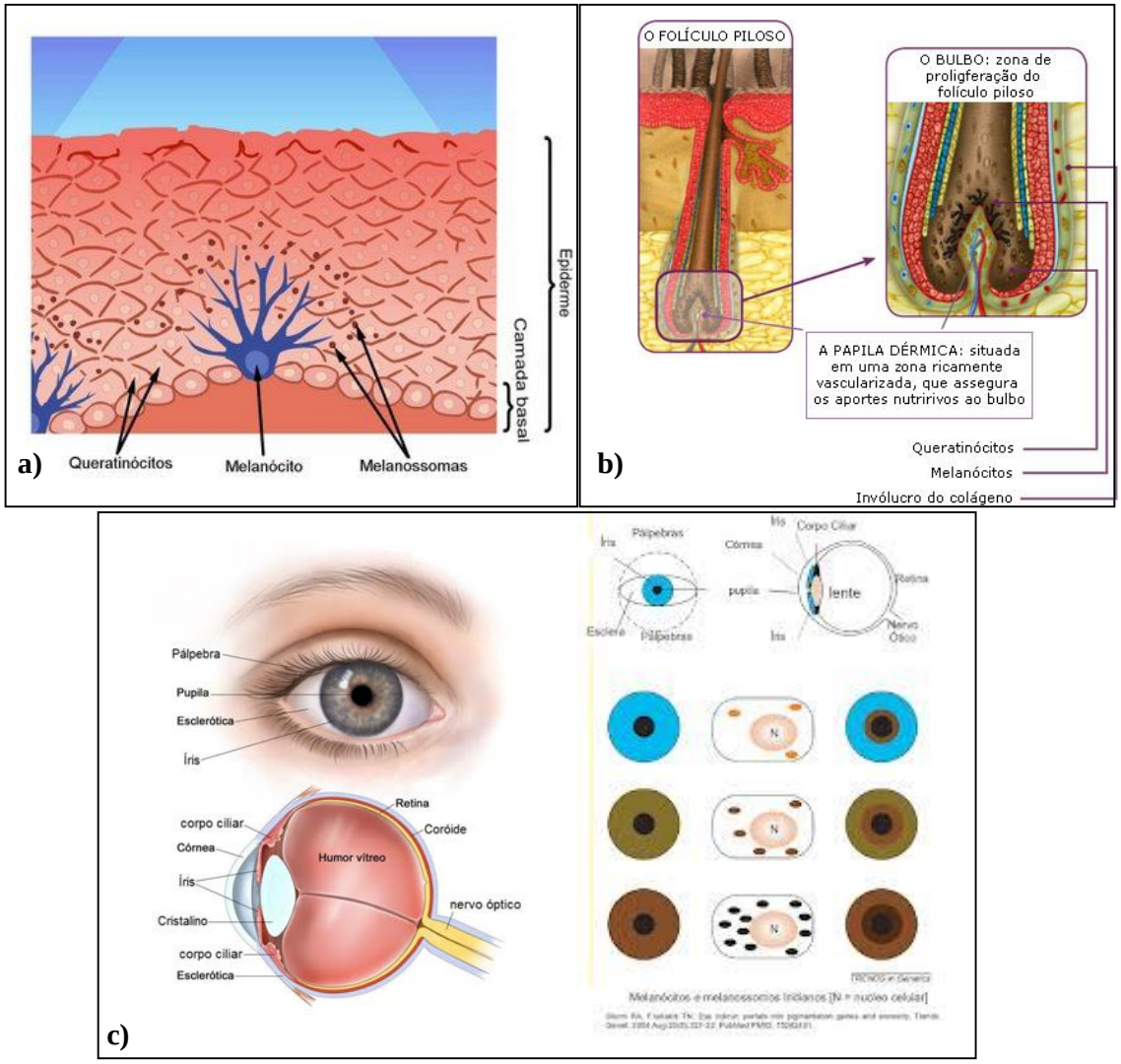


Figura 2. Regiões onde ocorre a produção de melanina mostrando as principais estruturas envolvidas neste processo. a) Corte da epiderme mostrando o melanócito e a liberação dos melanossomos na pele. b) Corte do folículo piloso mostrando o local do bulbo onde ocorre a produção de melanina. c) Região da íris no olho humano e tipos de deposição diferencial de melanina nos olhos com diferentes cores. Para mais detalhes, favor consultar texto.

A íris, por sua vez, possui duas camadas de tecido: a camada íntima, chamada de epitélio de pigmento da íris (do inglês *IPE*, *Iris Pigment Epithelium*), que não contribui para a variação da pigmentação normal da cor da íris, pois, apesar de ser altamente pigmentado, o conteúdo de melanina neste local possui uma distribuição similar em indivíduos com diferentes cores de íris (exceto albinos, que possuem falta de melanina e os olhos aparentam ser rosa como um resultado da reflexão da luz dos vasos sanguíneos); e a camada externa, que é conhecida como o estroma da íris, sendo o conteúdo de melanina nesta camada a causa primária da variação normal da cor da íris (Imesch *et al*., 1997; Sturm & Frudakis, 2004; Wielgus & Sarna, 2005; Sturm & Larsson, 2009). A íris castanha possui grandes quantidades de melanina e elevado número de melanossomos, absorvendo uma proporção substancial de luz incidente, particularmente de comprimento de onda curto. Ao contrário, a íris de cor azul possui pouco conteúdo de melanina e poucos melanossomos (Figura 2c). Como resultado, a luz com maior comprimento de onda penetra no estroma e é absorvida na camada íntima, enquanto a luz solar com comprimento de onda menor (azul) é refletida pela matriz de colágeno da íris, o que nos faz perceber visualmente a cor azul. A cor mel e verde possuem uma quantidade de melanina intermediária (Sturm & Frudakis, 2004).

1.5 Mecanismos moleculares da pigmentação

Os mecanismos fisiológicos descritos anteriormente e os correspondentes fatores genéticos subjacentes a estes já são conhecidos há pelo menos duas décadas, em especial pelo estudo de doenças que afetam a pigmentação como é o caso do albinismo (King *et al*., 1994). Entretanto, é mais recente o entendimento da variação normal da pigmentação humana (Makova & Norton, 2005). Pode-se dizer que o principal fator que alavancou o interesse nesta área foi a descoberta de mutações no gene do receptor de melanocortina 1 (*MC1R*), que são associadas com a variação normal da cor da pele (Valverde *et al*., 1995). Até 2002, somente o receptor de melanocortina 1 (*MC1R*) havia sido estudado em detalhes sobre a variação normal da cor da pele e cabelos em humanos (Rees, 2003), enquanto que em camundongos já se estimava um número acima de 125 *loci* responsáveis pela coloração da pelagem e outros fenótipos (Bennet & Lamoreux, 2003). No endereço eletrônico a seguir pode-se ter uma idéia do número atualizado de *loci* responsáveis pela coloração em camundongos: *http://www.espcr.org/micemut*.

O produto do gene *MC1R* é um receptor de superfície celular nos melanócitos e funciona principalmente na via de produção da melanina (Makova & Norton, 2005). Este produto regula a troca entre a produção de eumelanina e feomelanina (Barsh, 1996; Sturm *et al.*, 2001), que, como já vimos, constituem os dois principais pigmentos da camada basal da epiderme, do bulbo capilar e da íris (Makova & Norton, 2005; Parra, 2007). O receptor de melanocortina 1 pertence a uma família de receptores acoplados à proteína G. Nos mamíferos, este receptor responde aos hormônios alfa-melanócito-estimulante (α-MSH) e adrenocorticotrófico (ACTH), ambos produtos da pro-opiomelanocortina (*POMC*). Quando o *MC1R* é ativado pelo α-MSH, os níveis de cAMP (do inglês, *Ciclic Adenosine Monophosphate*) intracelulares são aumentados via ativação da adenilil ciclase, levando à produção de níveis elevados de tirosinase (produto do gene *TYR*) e, consequentemente, à produção de eumelanina (Makova & Norton, 2005).

De forma contrária, a proteína de sinalização antagonista agouti (produto do gene *ASIP*, que foi descrita primeiro em roedores, daí veio o nome) pode se ligar ao produto do *MC1R*, bloqueando a ativação pelo α-MSH. Quando o produto do gene *ASIP* bloqueia completamente essa ativação, a feomelanina é produzida no lugar da eumelanina (Kanetsky *et al.*, 2002). A Figura 3 mostra didaticamente o mecanismo descrito anteriormente. É importante lembrar que este mecanismo não é tão simples como parece. Os grânulos de melanina são sintetizados usando o aminoácido tirosina como o substrato principal. A tirosina também é um aminoácido precursor das catecolaminas - dopamina, adrenalina e noradrenalina (Fernstrom & Fernstrom, 2007); e isto pode ajudar a explicar a relação entre pigmentação humana e desenvolvimento cerebral mencionada anteriormente. A tirosinase, por sua vez, é uma enzima chave na melanogênese e permeia os primeiros passos na síntese de melanina, que envolve a hidroxilação da tirosina para dopa e a subsequente oxidação para dopaquinona (Parra, 2007). A dopaquinona, por sua vez, é um composto intermediário que passa por modificações adicionais em duas vias alternativas. Na ausência do aminoácido cisteína, a dopaquinona eventualmente origina eumelanina, e, na presença da cisteína, é produzida a feomelanina (Ito, 2003).

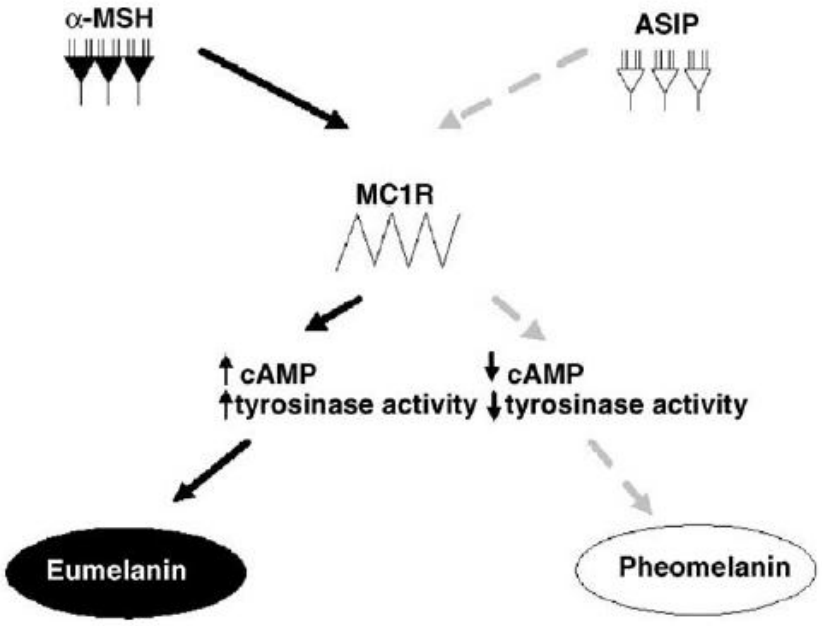


Figura 3. Diagrama dos efeitos alternativos da ligação do α-MSH ou *ASIP* ao receptor de melanocortina 1. A ativação pelo α-MSH leva à produção de eumelanina (castanho/preto), enquanto que a ligação do antagonista *ASIP* leva à produção da feomelanina (vermelha/amarela). Fonte: Makova & Norton, 2005.

Adicionalmente ao eixo *MC1R-TYR-ASIP* visto anteriormente, Yamaguchi & Hearing (2009) apresentam uma excelente visão geral de fatores fisiológicos diversos que podem influenciar na pigmentação humana (Figura 4). Além da tirosinase, uma enzima crítica dependente de cobre que, como vimos, é requerida para a síntese de melanina, os componentes enzimáticos dos melanossomos incluem a proteína relacionada à tirosinase 1 (produto do gene *TYRP1*), e dopacromo tautomerase (produto do gene *DCT*). Estas três enzimas cooperam entre si para sintetizar os dois tipos distintos de melanina (Costin *et al.*, 2005). Alguns fatores de transcrição importantes no processo de síntese de melanina também são dignos de nota. O fator de transcrição mais crítico que regula a função dos melanócitos é o fator transcricional microftalmia (produto do gene *MITF*) (Steingrimsson *et al.*, 2004). O promotor do gene *MITF* é regulado, por sua vez, por vários outros fatores de transcrição, incluindo os produtos dos genes *PAX3, SOX9* (Passeron *et al.*, 2007), *SOX10, LEF-1/TCF* e *CREB* (Lin & Fisher, 2007).

Além dos processos já vistos, o óxido nítrico, além do seu conhecido papel no relaxamento da musculatura lisa, também influencia o início da melanogênese, o eritema e a

imunossupressão em resposta à radiação UV. Além disso, alguns fatores hormonais que estimulam a pigmentação incluem o estrógeno (o que se relaciona com a pigmentação induzida pela gravidez), o ACTH, a endorfina, e o já mencionado α-MSH. Os andrógenos, por sua vez, possuem efeitos inibitórios nos melanócitos (Takodoro *et al.*, 2003). Modificações na pigmentação também são clinicamente observadas em resposta à inflamação. Mediadores químicos inflamatórios derivados do araquidonato, especialmente as prostaglandinas, os leucotrienos e os tromboxanos, são responsáveis pela pigmentação induzida pela inflamação, uma vez que são estimuladores da atividade do gene *TYR* (Costin & Hearing, 2007). É importante salientar que nem todas citocinas inflamatórias aumentam a inflamação. A IL-6 e o TNF-α são também conhecidos por suprimir a pigmentação da pele, o que sugere a importância de se elucidar a relação entre a melanogênese e inflamação.

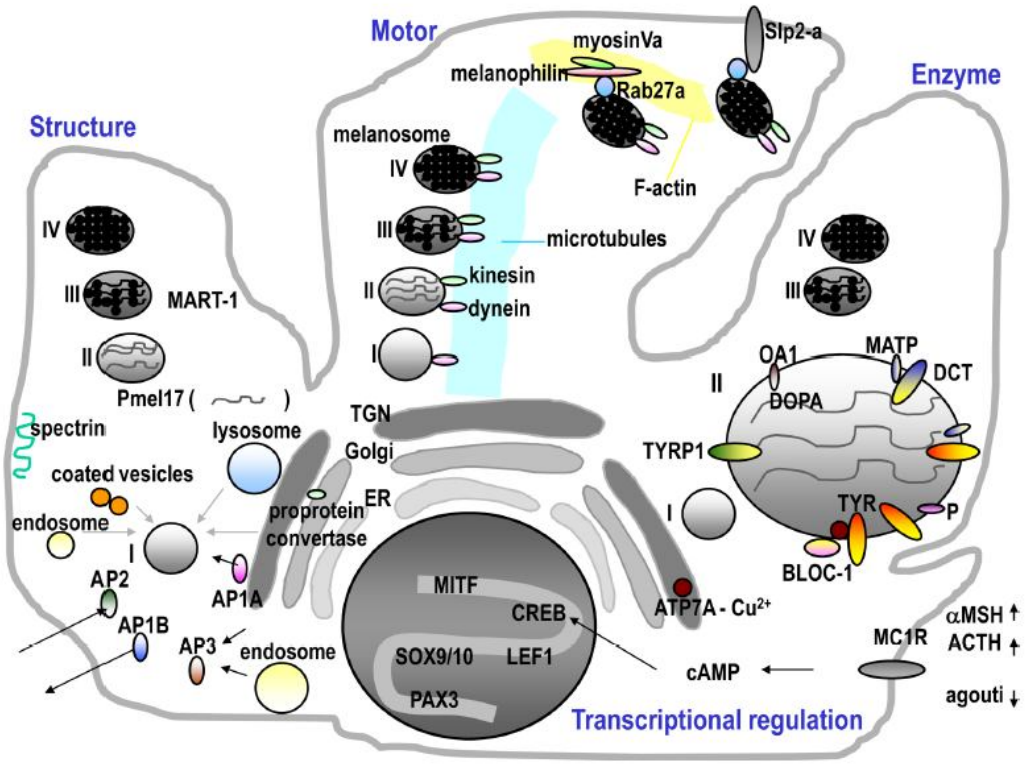


Figura 4. Fatores que afetam a pigmentação da pele dentro dos melanócitos. Este esquema apresenta os componentes envolvidos na estrutura dos melanossomos (esquerda), transporte (centro), enzimas (direita), bem como os fatores transcricionais relacionados (margem inferior). Fonte: Yamaguchi & Hearing (2009).

Enfim, compreender o contexto biológico da produção da melanina é uma necessidade para o bom entendimento da rede da pigmentação, mas detalhar neste tópico a função de cada um dos genes responsáveis direta ou indiretamente pela síntese, armazenamento, transporte e outros processos relacionados se tornaria uma tarefa extretamente extenuante. Desta forma, informações detalhadas sobre a localização, mutações e função de vários dos genes envolvidos nestes processos podem ser encontradas em Cerqueira *et al.* (2011; Apêndice II), bem como em Sulem *et al*. (2007), Yamaguchi & Hearing (2009) e Sturm & Duffy (2012) e nas referências citadas nestes trabalhos.

O que também é importante mencionar aqui é que existem atualmente alguns genes que são considerados chave em Europeus e que são responsáveis pela variação de cor de olhos e cabelos, que são as duas características mais estudadas em pigmentação humana. Por exemplo, três polimorfismos no *MC1R* estudados na presente tese (rs1805009, rs1805008, rs1805007) são clássicos na via da pigmentação, sendo comumente associados com cor de olhos e cabelos (Smith *et al.*, 1998; Sulem *et al*., 2007, 2008; Latreille *et al*., 2009), embora acredita-se que eles também respondam por uma substancial variação da cor da pele (Duffy *et al*., 2004; Rees, 2004; Makova & Norton, 2005; Sulem *et al.*, 2008). Estes três SNPs possuem uma característica muito importante: estão entre os principais preditores de cor de cabelos ruivos em populações Européias (Duffy *et al*., 2004). Estes três polimorfismos são conhecidos há algum tempo como "RHC" (*Red hair color*) (Box *et al*., 1997; Sturm *et al*., 2001).

Com relação à cor dos olhos, existem evidências sugerindo que o gene *HERC2* influencia a expressão do *OCA2* (Eiberg *et al*., 2008; Kayser *et al*., 2008; Sturm *et al*., 2008; Branicki *et al*., 2009; Donnelly *et al*., 2012; Visser *et al*., 2012). Uma análise realizada por Sturm & Larsson (2009), sobre uma seqüência de 11 pares de bases altamente conservada (TGACA[T/C]TTAAT), contendo o polimorfismo rs12913832 localizado centralmente, sugeriu que este poderia representar um sítio de ligação consenso para o fator de transcrição tipo-helicase (HTLF) (Sturm *et al.*, 2008). Criticamente, esta seqüência consenso é abolida no alelo rs12913832*C, que confere olho azul. Os autores propuseram um modelo de como este mecanismo regulatório atua na determinação da cor do olho. Quando ativo, o alelo rs12913832*T apresenta o reconhecimento para o *HTLF*, o que promove o desenrolamento da cromatina, expondo sequências regulatórias adicionais para fatores de transcrição como o *MITF* e *LEF1*, que, por sua vez, levam à expressão do gene *OCA2*. A proteína Oca2 então

atua na via de maturação do melanossomo para produzir uma íris castanha amplamente pigmentada. O alelo rs12913832*C, por sua vez, leva a uma anulação da ligação do HTLF, fazendo com que a cromatina permaneça compactada e indisponível para transcrição do lócus *OCA2*. Como uma consequência da perda seletiva da produção do *OCA2* nos melanócitos da íris, a inabilidade em produzir melanossomos maduros resulta na aparência de uma íris azul. Neste modelo, a falha em ativar a expressão do gene *OCA2* é a explicação molecular mais parcimoniosa existente. De fato, um aumento significante no nível de mRNA transcrito do gene *OCA2* foi observado no alelo rs12913832*T (olho castanho), comparado com o alelo em homozigose rs12913832*C (olho azul) (Cook *et al*., 2009), mostrando a consistência da afirmação de que o elemento conservado atua dentro de uma região controladora de lócus. Embora os estudos sobre a genética da cor de olhos têm se focado em polimorfismos *OCA2-HERC2* em populações Européias, o estudo de Iida *et al*. (2009), compreendendo 523 Japoneses, verificou que esta população parece não possuir o alelo rs12913832*C, consistente com a ausência de cor de olho azul nestes indivíduos.

1.6 Polimorfismos genéticos relacionados com a pigmentação e sua relação com a história demográfica humana

Os dados genéticos disponíveis para alguns genes candidatos à pigmentação (por exemplo, o *MC1R*) são consistentes com a hipótese da seleção natural favorecendo a pigmentação escura em áreas tropicais. O gene *MC1R* é, sem dúvida, o mais exaustivamente estudado como candidato à pigmentação. O papel deste gene na pigmentação normal foi primeiro mostrado por Valverde *et al*. (1995), que apresentaram uma associação de variantes do *MC1R* com cabelos ruivos e pele clara. O gene *MC1R* apresenta um interessante padrão de polimorfismos ao redor do mundo (Makova & Norton, 2005). O quadro geral de polimorfismos é consistente com a ação da seleção purificadora removendo mutações no *MC1R* que poderiam promover a síntese de feomelanina em indivíduos da região da África Sub-saariana, e possivelmente de outras regiões com altas incidências de UVR. A situação é bastante diferente na Europa, bem como no Leste e Sul da Ásia, onde o *MC1R* é altamente polimórfico. De fato, os níveis de diversidade nucleotídica observada no gene *MC1R* nestas populações são mais elevados que os valores observados em outros genes autossômicos (Makova & Norton, 2005). Mais de 20 variantes não-sinônimas têm sido relatadas em

populações européias (Sturm *et al.*, 2001). Algumas destas mutações (rs2228479 - Val92Met) promovem uma habilidade reduzida na ligação do *MC1R* com α-MSH, ou em ativar a adenilil ciclase (rs1805007 - Arg151Cys, rs1805008 - Arg160Trp, e rs1805009 - Asp294His), explicando a associação destas variantes com os fenótipos de pele e cabelos relacionados com altos níveis de feomelanina (Xu *et al.*, 1996; Frandberg *et al.*, 1998; Schioth *et al*., 1999).

Segundo Rana *et al.* (1999), o alto nível de polimorfismos do *MC1R* observado em europeus sustenta, em parte, a hipótese da vitamina D sugerida por Loomis (1967), Cavalli-Sforza *et al.* (1994), bem como outros autores, onde alelos diferentes seriam selecionados para maximizar a síntese de vitamina D, dependendo da exposição à UVR que os indivíduos em consideração estariam submetidos. Entretanto, segundo Harding *et al.* (2000), a alta quantidade de polimorfismos no lócus do *MC1R* em europeus poderia ser devido ao relaxamento de restrições funcionais. Eles aplicaram vários testes de neutralidade e não encontraram nenhuma evidência estatística de que a diversidade observada no *MC1R* tenha sido intensificada pela seleção. Portanto, os dados sobre os alelos do *MC1R* em populações européias são consistentes com a remoção das restrições funcionais ao invés da hipótese de seleção em favor das mutações de perda de função. O que se sabe é que a região codificante do gene *MC1R* em asiáticos e africanos é menos polimórfica em comparação com europeus (Harding *et al*., 2000; Rana *et al*., 1999; Przeworski *et al*., 2000).

Outro interessante estudo sobre as frequências dos marcadores genéticos de pigmentação e a história da dispersão do homem no planeta foi recentemente publicada por Beleza *et al*. (2012). Este estudo mostrou que não só o *MC1R*, como também outros genes relacionados com pigmentação que eles estudaram possuem um padrão geoespacial de variação genética que são correlacionadas com a distribuição geográfica da cor da pele e mostram uma forte evidência de varredura seletiva em populações européias e asiáticas (Izagirre *et al*., 2006; McEvoy *et al*., 2006; Soejima *et al*., 2006; Voight *et al*., 2006; Lao *et al*., 2007; Myles *et al*., 2007; Norton *et al*., 2007; Sabeti *et al*., 2007; Williamson *et al*., 2007; Pickrell *et al*., 2009). Além disso, o quadro geral formado pelos dados genéticos sugere que a pigmentação da pele tem uma história evolutiva complexa, com a seleção natural atuando sobre diferentes genes, em diferentes momentos, e em distintas populações ao longo dos continentes (McEvoy *et al*., 2006; Lao *et al.*, 2007; Myles *et al*., 2007; Norton *et al*., 2007; Pickrell *et al*., 2009).

Alguns genes, como o *KITLG*, possuem alelos derivados (associados com pigmentação clara) que alcançam altas frequências tanto em europeus quanto em populações do Leste Asiático e provavelmente foram selecionados em uma população proto-eurasiana (Lao *et al.*, 2007; Williamson *et al.*, 2007; Pickrell *et al.*, 2009). Outros genes possuem alelos derivados com uma distribuição principalmente restrita à Europa (*SCL24A5*, *SLC45A2* e *TYRP1*) ou ao leste da Ásia (*DCT* e *ATRN*), sugerindo que genes diferentes são responsáveis pela redução do conteúdo da melanina em europeus ou asiáticos (McEvoy *et al.*, 2006; Lao *et al.*, 2007; Myles *et al.*, 2007; Norton *et al.*, 2007; Pickrell *et al.*, 2009). O reconhecimento destes padrões reforça ainda mais a questão da convergência evolutiva para pigmentação humana. Evidências adicionais de adaptação convergente são fornecidas por estudos em genes como o *OCA2*, no qual diferentes alelos de branqueamento da pele têm sido provavelmente submetidos a independentes varreduras em europeus e asiáticos (Lao *et al.* 2007; Edwards *et al.* 2010; Donnelly *et al.*, 2012).

1.7 Como definir a pigmentação de pele, olhos e cabelos

Há consenso na literatura científica de que é extremamente necessário refinar a determinação do fenótipo da pigmentação de uma categoria qualitativa (subjetiva e muito sujeita a erros) para uma medida mais quantitativa do conteúdo da melanina (Makova & Norton, 2005), pois categorizações subjetivas da cor podem inibir a reprodutibilidade dos experimentos e a comparação direta dos resultados com outros estudos. Nos estudos de pigmentação da pele é usual utilizar a classificação de Fitzpatrick *et al.* (1987) como medida do conteúdo de melanina (Valverde *et al.*, 1995; Smith *et al.*, 1998; Healy *et al.*, 2000; Palmer *et al.*, 2000; Dimisianos *et al.*, 2009; Neitzke-Montinelli *et al.*, 2012). Esta escala classifica a pele humana em quatro tipos e se baseia na resposta à incidência solar que os indivíduos possuem. A pele do tipo I é aquela que sempre se queima, mas nunca se bronzeia; a pele do tipo II é aquela que se queima, mas também se bronzeia levemente (bronzeamento suave); no tipo III é classificado aquele indivíduo que algumas vezes se queima sob o sol, mas sempre se bronzeia (bronzeamento moderado); e, por fim, o tipo IV, que nunca se queima, e sempre fica bronzeado (bronzeamento intenso) (Fitzpatrick *et al.*, 1987). Pode-se observar que esta escala é bastante subjetiva e, por isso, formas quantitativas e mais objetivas foram desenvolvidas ao longo dos anos. Alguns avanços tecnológicos têm permitido medir facilmente a melanina na pele ou cabelos usando tecnologias de reflectância, como, por

exemplo, a reflectometria triestímulos, espectroscopia de banda estreita e espectroscopia de reflectância difusa (Shriver & Parra, 2000; Parra, 2007).

No estudo do Shriver & Parra (2000) foram utilizados dois reflectômetros de mão computadorizados: o colorímetro *Photovolt ColorWalk* (*Photovolt, UMM Electronics, Indianapolis*) e o *DermaSpectrometer* (um reflectômetro especializado em banda estreita – *narrow-band*; *Cortex Technology, Hadsund, Denmark*) para comparar dois métodos de determinação objetiva para cor de pele e cabelos, ambos instrumentos estão disponíveis pelo menos há 50 anos no mercado e são comumente utilizados na dermatologia, antropologia e biologia. Estes instrumentos determinam a cor medindo a intensidade de luz refletida num comprimento de onda particular. A diferença entre ambos os aparelhos é que o *Photovolt ColorWalk* emite uma luz branca (fotodiodo) para medir a intensidade da luz refletida. Já o *DermaSpectrometer*, no lugar de luz branca e fotodiodos, usa dois LEDs (diodos emissores de luz), um verde e outro vermelho, iluminando uma superfície, e então registrando a intensidade da luz refletida. O resultado das medidas pode então ser expresso por vários sistemas padronizados de cor, sendo considerado o mais importante o sistema CIELab (*Commission International d'Eclairage* Lab), estabelecido em 1976, no qual qualquer cor pode ser descrita por três valores: L*, a luminosidade; a*, a quantidade de verde ou vermelho; e b*, a quantidade de amarelo ou azul. O resultado da leitura dos aparelhos também pode ser visto em termos de índices de eritema e melanina (E & M) ou ainda, por medida de RGB (*Red-Green-Blue*).

No estudo do Shriver & Parra (2000) foi concluído que ambos os tipos de instrumentos fornecem uma acurada estimativa do nível de pigmento na pele e cabelos; entretanto, as medidas utilizando o *DermaSpectrometer* podem ser menos afetadas por grande vermelhidão em certas regiões do corpo devido a vascularização aumentada. Além disso, os autores observaram uma alta correspondência entre a medida L* e o IM. No *link* a seguir é possível transformar valores de vários sistemas de medidas de cor entre si através de uma matemática simples: *http://www.easyrgb.com/index.php?X=MATH*. Com recentes avanços técnicos na área da colorimetria e fotometria, novos instrumentos têm sido criados, oferecendo vantagens substanciais (precisão, portabilidade e facilidade de uso) sobre instrumentos prévios (Parra, 2007). Um desses instrumentos é o novo *DermaSpectrometer II* (Figura 5), que possui um *software* interno para registrar valores de diferentes sistemas de cor simultaneamente, o qual

foi utilizado para o presente trabalho com a finalidade de mensurar exclusivamente a cor da pele.

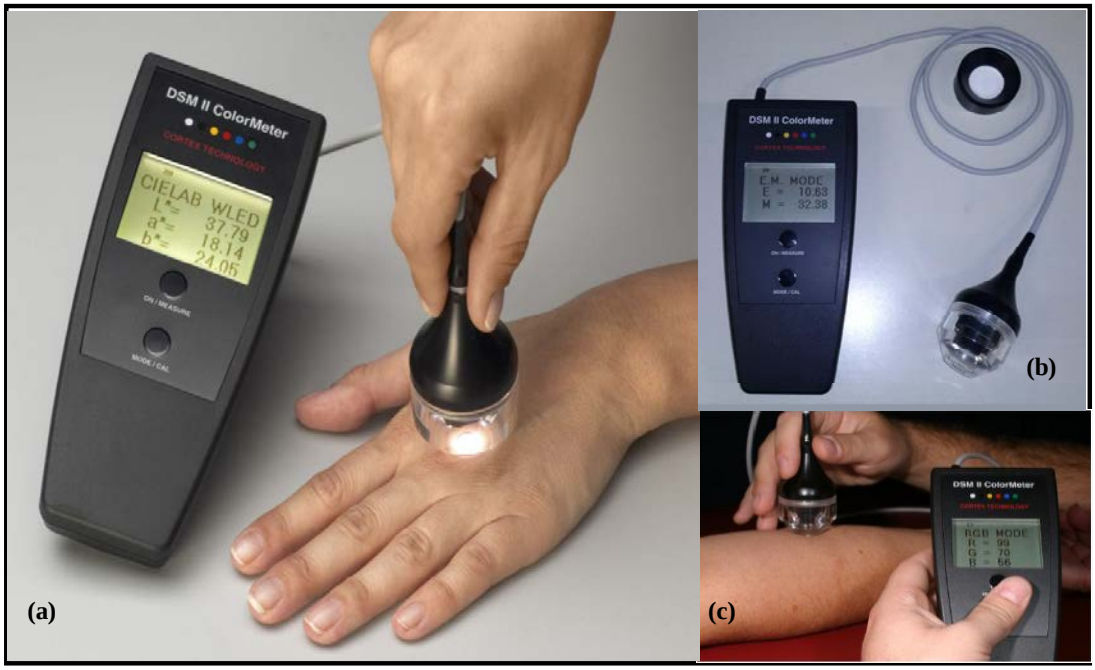


Figura 5. Fotografias do *DermaSpectrometer II (DSM II) ColorMeter*. Como mostrado nas fotos, este aparelho oferece uma medida de cor em três sistemas simultaneamente: a) CIELab; b) E&M (Eritema e Melanina); c) RGB. Suas configurações padrões foram criadas principalmente para aplicações relacionadas com a pele. Maiores informações, vide texto. Fonte: *http://www.micglobal.co.uk/derma-spectrometer-i968.html* e arquivo pessoal de fotos.

É importante mencionar aqui que o protocolo de coleta de dados do consórcio CANDELA envolveu a obtenção de medidas quantitativas de pigmentação da pele através dos índices de melanina em regiões não-expostas ao sol (região inferior do braço com menor incidência solar) como descrita e utilizada por Stokowski *et al*. (2007) e Shriver & Parra (2000). Deste modo, as medidas foram obtidas pelo espectrofotômetro de reflectância, conhecido também como *DSM II ColorMeter* (Figura 5). Este instrumento emite luz no comprimento de onda verde (568nm) e vermelho (655nm) do espectro visível e o foto-detector mede a quantidade de luz refletida pela pele. Estas medidas são utilizadas para se estimar o conteúdo de melanina da pele, a qual é expressa como E&M (Eritema e Melanina). Em populações humanas o IM varia de cerca de 20 (indivíduos de pele branca) até próximo de 100 (indivíduos com pele escura). Na busca de capturar uma leitura mais apurada da

pigmentação de pele, estas medidas foram preferencialmente realizadas em ambientes fechados e bem iluminados.

Com relação à cor dos olhos (íris), vários estudos já tentaram estabelecer um sistema universal de classificação para esta característica (para uma revisão mais detalhada, consulte Mackey *et al.*, 2011). O antropólogo suíço Rudolf Martin (Universidade de Zurich) foi quem produziu a primeira tentativa de se usar um conjunto padronizado de olhos artificiais para classificação da cor da íris. Em 1903 ele desenvolveu o painel da cor dos olhos, que consistiu em 16 olhos progressivamente arranjados em tons do castanho escuro (número 1) ao azul claro (número 16) (Figura 6; Oetteking, 1926). Vários estudos de genética também têm se baseado em classificações qualitativas (i.e. azul/cinza, verde, mel, castanho claro, castanho escuro/preto) (Eiberg & Mohr, 1996; Rebbeck *et al.*, 2002; Zhu *et al.*, 2004; Duffy *et al.*, 2007). Esta classificação pode ser falha em capturar a natureza quantitativa da variação da pigmentação da íris, como mencionado. Além disso, vários fatores podem modificar a cor da íris, como exemplo: a cor dos olhos pode mudar com diferentes níveis de dilatação da pupila; imagens de fotografias desfocadas tendem a resultar numa cor de olhos mais uniforme (perdendo a natureza heterogênea da cor da íris); a maquiagem, no caso das mulheres, pode modificar sensivelmente a impressão do observador de um olho castanho claro para um castanho escuro e vice-versa (Mackey *et al.*, 2011). Existem alguns esforços atualmente no desenvolvimento de métodos de quantificação a cor da íris baseada na análise de fotografias, tomadas sob condições padronizadas. Como exemplo, German *et al*. (1998) e Niggermann *et al.* (2003) relataram a pigmentação da íris no espaço de cor RGB; Melgosa *et al*. (2000) quantificaram a cor da íris usando o sistema CIELab; o estudo de Fan *et al.* (2003), por sua vez, desenvolveu um método que extrai automaticamente a região da íris na fotografia, computa a cor da íris, e corrige a cor baseada em um cartão de calibração padrão; e o estudo de Frudakis *et al.* (2007) tentou utilizar espectroscopia digital para predizer acuradamente o conteúdo de melanina da íris.

Mais recentemente, um trabalho que, segundo Ruiz *et al*. (2013), se destacou pela robustez das medidas foi o estudo feito por Liu *et al*. (2010). Após comparar vários sistemas de cor para representar a cor da íris, incluindo o RGB, CIELab, CIEXYZ (este último é denominado de valor triestímulos, e se baseia numa representação gráfica tridimensional, onde cada eixo refere-se a uma cor primária – verde, vermelho e azul) e HSV/HSB (*Hue* – matiz, *Saturation* – saturação, *Value* = valor, *Brightness* - brilho), os autores escolheram

representar a cor da íris por este último sistema (HSV/HSB). Neste trabalho, foram tomadas fotografias dos olhos dos voluntários numa resolução de 800x600 pixels, os valores de RGB na região da íris de cada imagem foram anotados, e, posteriormente, os valores de RGB foram transformados para valores HS de acordo com fórmulas-padrão. Segundo Liu *et al.* (2010), este sistema se ajusta bem na aplicação de estudos genéticos porque (1) os valores H e S da escala são independentes da luminosidade (dada pelo valor B), que varia bastante dependendo do local onde a foto foi tomada (o valor B pode ser excluído na análise genética); (2) os valores de H e S podem representar o tipo e a quantidade de pigmento da íris; e (3) os valores de H e S podem ser diretamente traduzidos para outros sistemas de cores (para saber mais sobre a matemática envolvida na conversão das medidas, acesse *http://www.easyrgb.com/index.php?X=MATH*).

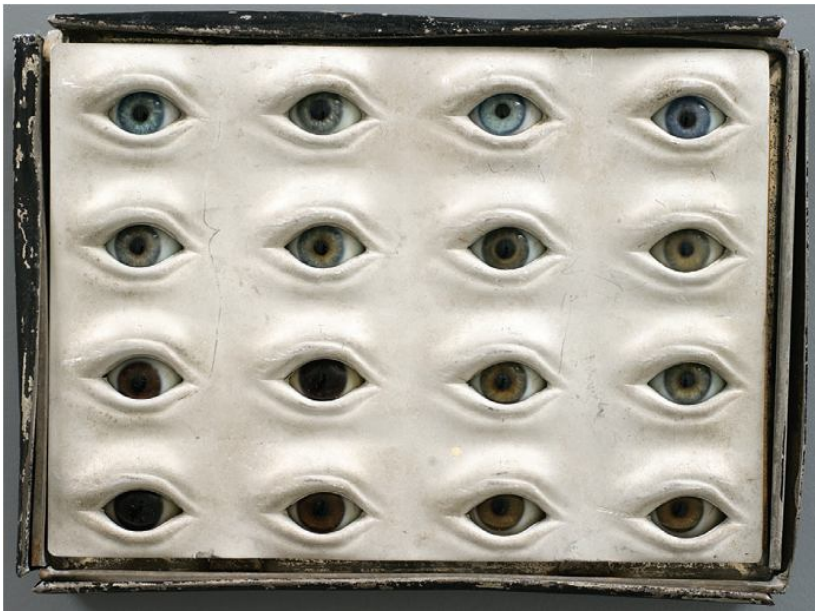


Figura 6. Painel representando os 16 olhos de Rudolf Martin, classificados do castanho mais escuro (número 1) ao azul mais claro (número 16). Cabe salientar que os olhos neste quadro não seguem uma ordem definida. Fonte: Mackey *et al.* (2011).

Embora os métodos de classificação fenotípica da íris por análise quantitativa através de fotografias digitais têm sido propostos (Melgosa *et al*. 2000; Niggermann *et al.* 2003; Liu *et al.*, 2010), a caracterização fidedigna e a designação de fenótipos intermediários da cor da íris permanece difícil e ainda relativamente subjetiva (Frudakis *et al*., 2007; Ruiz *et al*., 2013). Um ponto importante a salientar é que a íris humana também possui outras características que não são mensurados através da quantificação utilizada para verificar a cor

dos olhos e estes fatores também estão sob forte influência genética (Larsson *et al*., 2003). Para agravar ainda mais a situação, sabe-se que embora a cor dos olhos seja assumida como fixa por toda vida adulta, existem mudanças com a idade ou mudanças no estado patológico do indivíduo que podem afetar esta característica (Stjernschantz *et al*., 2002).

Com relação à qualificação da cor dos cabelos e dos olhos dos voluntários do presente trabalho, uma ficha fenotípica do consórcio CANDELA foi utilizada (Anexo IX). Esta ficha foi preenchida pelo investigador no momento da coleta, indicando a cor dos cabelos em uma das 4 categorias, a saber: ruivo/avermelhado, loiro, loiro escuro/castanho claro, e castanho escuro/preto; e a cor dos olhos, por sua vez, em uma das 5 categorias, são elas: azul/cinza, verde, mel, castanho claro, e castanho escuro/preto. Os fenótipos de cor dos olhos e cabelos foram marcados mediante confirmação dada por cada um dos voluntários da pesquisa. Categorias similares são comumente utilizadas na literatura (Sulem *et al*., 2007; Eiberg *et al*., 2008). Um esforço para quantificar a cor dos olhos e cabelos dos voluntários do CANDELA a partir das fotos está em andamento, baseado em protocolos mais recentes e precisos (Liu *et al*., 2010). A descrição suscinta da metodologia de quantificação está descrita no Apêndice I.

Como se observa, a classificação objetiva da pigmentação é extremamente importante para um estudo genético de associação. A falta desta objetividade pode trazer resultados espúrios em pesquisas desta natureza. Por exemplo, no estudo realizado por Bastos *et al*. (2009) uma comparação entre a auto-classificação e a hetero-classificação (classificação atribuída por outro) de indivíduos dentro de categorias de "cor/raça" indicou que pessoas com mais de 40 anos de idade tiveram uma chance 2,1 vezes maior de se classificarem como pardos do que como brancos, quando abordados por entrevistadoras negras, em relação às brancas. Os resultados deste estudo indicaram o caráter complexo da classificação qualitativa, sugerindo a influência de características pessoais das entrevistadoras sobre a auto-classificação de "cor/raça" dos entrevistados. Na tentativa de alertar sobre a padronização na classificação de pigmentação para os estudos genéticos que estão sendo realizados, o capítulo de livro publicado por nós (consultar Cerqueira *et al*., 2011 – Apêndice II) discorre sobre o assunto, trazendo ainda outras informações pertinentes na área. Na presente tese, durante o preenchimento das fichas fenotípicas dos voluntários, houve também o cuidado de fazê-lo mediante a observação de dois pesquisadores membros do consórcio CANDELA, para que as marcações fossem feitas de forma mais objetiva, no sentido de minimizar as variações oriundas das impressões pessoais na análise. No caso de discordância entre os dois

observadores, uma terceira opinião foi requerida. Isso foi feito para evitar desvios na hetero-classificação qualitativa de algumas características, como cor de cabelos e cor dos olhos.

1.8 Identificação humana pelo DNA e aplicações forenses dos estudos de pigmentação

A ciência molecular forense tem sido dominada pelas pesquisas em nível de DNA nas últimas décadas. A razão para este fato é o grande sucesso do uso de tecnologias relacionadas na pratica forense, o que tem revolucionado a polícia moderna (Bauer, 2007). O princípio básico destas pesquisas é que, exceto para gêmeos monozigóticos, cada ser humano é geneticamente diferente e único. Desta forma, qualquer amostra biológica contendo DNA, quando examinada em detalhes suficientes, revela que foi originada de somente um indivíduo específico (Pena *et al*., 1995). Alguns exemplos das vantagens oferecidas pela individualidade genética são: identificação de vítimas de crimes ou de acidentes de massa ou catástrofes naturais, identificação de criminosos por vestígios em cenas de crime, investigação de paternidade, bem como estudos em evolução humana, dentre outras aplicações.

O repertório de marcadores genéticos utilizados nas rotinas forenses tem crescido substancialmente e vários avanços nesta área têm trazido progresso para as ciências forenses nas últimas três décadas. O primeiro método genético usado para identificação humana utilizou polimorfismos de comprimento de fragmentos de restrição (*RFLPs*, do inglês, *Restriction Fragment Length Polymorphisms*) para tipagem de números variáveis de repetições consecutivas (*VNTRs*, do inglês, *Variable Number of Tandem Repeats*), que são fragmentos de DNA de 6-100 pares de bases que se repetem um atrás do outro por centenas de vezes (Wyman & White, 1980; Jeffreys *et al*., 1985; Budowle & Baechtel, 1990; Goodwin *et al*., 2011). Pouco tempo depois, os VNTRs foram substituídos pela análise de repetições curtas consecutivas (*STRs*, do inglês, *Short Tandem Repeats*) (Edwards *et al*., 1991; Budowle *et al*., 1998; Budowle *et al*., 2001), que são fragmentos de DNA de 1-6 pares de bases que se repetem *in tandem* centenas a milhares de vezes, sendo este atualmente utilizado por bancos de dados criminais e cíveis de DNA em todo o mundo (Jobim *et al*., 2006; Budowle & Van Daal, 2008; Goodwin *et al*., 2011). Os STRs possuem algumas vantagens em relação aos VNTRs. Por exemplo, o tamanho menor do fragmento amplificado e a capacidade de amplificação através da reação em cadeia da polimerase (*PCR*, do inglês, *Polymerase Chain*

Reaction), de amostras com relativo nível de degradação (Budowle & Van Daal, 2008; Goodwin *et al*., 2011). Para mais detalhes a respeito dos marcadores polimórficos para identificação humana, consultar Jobim *et al.* (2006) e Goodwin *et al*. (2011).

Além dos STRs, análises com os polimorfismos de base única comumente bi-alélicos (SNPs) também estão sendo utilizadas em testes de identidade genética, seja através da informação sobre linhagens do mtDNA e do cromossomo Y, seja através de marcadores informativos de ancestralidade, ou ainda para a predição de fenótipos (Budowle & Van Daal, 2008). Porém uma das principais desvantagens no uso dos SNPs é o seu baixo poder de discriminação nos testes de identidade genética (Goodwin *et al.*, 2011). Para se ter uma idéia, na tipificação com uso de VNTRs, utilizava-se cerca de 6 *loci*, em média, com um excelente poder de discriminação entre dois indivíduos quaisquer; na tipificação por STRs utiliza-se de 9 a 22 *loci*; estima-se que, para se obter o mesmo poder de discriminação, uma análise com pelo menos 50-100 SNPs é necessária (Budowle & Van Daal, 2008; Goodwin *et al*., 2011). Entretanto, os SNPs também apresentam algumas vantagens: é necessário um fragmento de apenas 60-80 pares de bases para amplificação por PCR, enquanto que, para análise de STRs, é necessário um fragmento de cerca de 350 pares de bases (Whitaker *et al.*, 1995; Ballantyne, 1997). Isso é extremamente importante para amostras altamente degradadas provenientes de desastres em massa e catástrofes naturais. Além disso, a quantidade de DNA necessária para amplificar um SNP é da ordem de 100 picogramas; para os STRs, é necessário cerca de 0,5-1 nanograma de DNA; e os VNTRs necessitam de 10-25 nanogramas de DNA para amplificação de fragmentos de até 10.000 pares de bases (Giusti & Budowle, 1995). Excelentes discussões sobre o poder dos SNPs e a possível substituição dos STRs por eles podem ser encontrados em Chakraborty *et al.* (1999), Budowle (2004), Gill *et al.* (2004), Pakstis *et al*. (2010) e Schneider (2012). Além da identificação humana, Kayser & Knijff (2011) também fizeram uma revisão bastante interessante sobre a utilidade de ferramentas genéticas em casos criminais.

Outro aspecto que merece destaque no contexto da presente tese é que, atualmente, o primeiro passo na análise de amostra de DNA encontrada em cena de crime ou local de desastres naturais é a produção de um perfil de STRs utilizando *kits* de amplificação multiplex para essas amostras. Quando o perfil de STRs não se pareia com um banco de dados criminais e cíveis, quaisquer informações adicionais são extremamente valiosas (Werrett, 2005; Rohlfs *et al.*, 2012), incluindo testes de predição de fenótipos específicos

(Tully, 2007). Em outras palavras, nos casos em que as investigações policiais falham em identificar os agressores, e principalmente quando os vestígios biológicos encontrados em cenas de crime são de fontes desconhecidas àqueles existentes nos bancos de dados policiais, se espera que, num futuro próximo, haja uma confiabilidade na predição de características visíveis externamente (comumente referidas pela sigla *EVCs*, do inglês, *Externally visible characteristics*), com a tecnologia conhecida como Fenotipagem Forense por DNA (do inglês, *FDP – Forensic DNA Phenotyping*) ou inteligência pelo DNA (Kayser & Schneider, 2009; Kayser & Knijff, 2011).

Como visto anteriormente, a variação normal e patológica da pigmentação humana tem sido há muito tempo alvo de estudos médicos (Valverde *et al*., 1995; Duffy *et al*., 2004) e antropológicos (Jablonski & Chaplin, 2000; Parra *et al*., 2004) e, por último, objeto de investigação que visa a aplicação forense (Sulem *et al.*, 2007; Valenzuela *et al*., 2010). Sendo assim, nos últimos anos inúmeros trabalhos têm almejado identificar genes (e seus alelos variantes) responsáveis por características fenotípicas visíveis. Alelos comuns associados com a diversidade da pigmentação normal em humanos foram identificados em vários genes. Algumas variantes no gene *MC1R*, como já salientado em itens anteriores, estão associadas com cabelos vermelhos, peles claras, sardas, e câncer de pele (Valverde *et al*., 1995; Rees, 2004; Makova & Norton, 2005). Outros estudos relacionados com a cor de cabelo e olhos, por sua vez, têm revelado forte ligação da cor azul-castanha a uma região no cromossomo 15 que inclui o gene *OCA2* (codifica a proteína óculo-cutâneo II) (Eiberg & Mohr, 1987; Eiberg & Mohr, 1996; Posthuma *et al*., 2006), que também já havia sido ligado previamente com o albinismo. Esses dois exemplos ilustram uma pequena porção do que vem sendo descrito nesta esfera de interesse, cujos resultados já estão sendo utilizados em pequena escala no nível forense. Portanto, a habilidade de se predizer características físicas de um indivíduo diretamente a partir do material encontrado em uma cena de crime é algo bastante promissor.

Acredita-se que de todas as nossas características visíveis externamente (além do sexo), os traços de pigmentação, particularmente cor dos olhos e cabelos, sejam os mais promissores para predição de fenótipos com uso do DNA (Branicki *et al*., 2007; Branicki *et al*., 2009; Mengel-From *et al*., 2009; Liu *et al*., 2009; Mengel-From *et al*., 2010; Walsh *et al*., 2011a; Walsh *et al*., 2011b; Branicki *et al*., 2011; Spichenok *et al*., 2011; Walsh *et al*., 2012; Ruiz *et al*., 2013; Draus-Barini *et al*., 2013). Em 2001 ocorreu a publicação do primeiro exemplo de

um teste de predição de fenótipo para uso da inteligência policial em investigações criminais (Grimes *et al.*, 2001). O estudo descreve um protocolo de minisequenciamento multiplex para varredura rápida de amostras de DNA para presença de 12 variantes do gene *MC1R*. Numa outra investigação realizada por Frudakis *et al*. (2003), com 754 SNPs em 851 indivíduos, foram identificados 61 SNPs associados com cor da íris. Cerca de metade desses SNPs foram independentemente associados com cor da íris, enquanto o restante foi associado somente se considerados como parte de um haplótipo ou diplótipo (par diplóide do haplótipo). Em 2008 foi lançado o livro "*Molecular Photofitting*", que descreve o uso de SNPs para inferência de ancestralidade geográfica e predição de fenótipo no uso da perícia policial e algumas das análises realizadas para tal fim (Frudakis, 2008). Outros trabalhos têm sido publicados investigando valores preditivos de várias combinações de alelos de SNPs no gene *OCA2*, bem como outros (*HERC2*, *SLC45A2*/*MATP*) para cor de olho azul/castanho (Mengel-From *et al.*, 2010). Detalhes adicionais sobre uso de marcadores genéticos de pigmentação para aplicação forense podem ser encontrados em Gehring (2010) e Souza (2012).

Recentemente, pesquisadores do Instituto Erasmus na Holanda, liderados pelo Prof. Manfred Kayser, fizeram com que a expectativa de se predizer traços complexos como a pigmentação dos olhos e cabelos se tornasse mais real após a publicação do estudo de Walsh *et al.* (2013). Neste estudo foi desenvolvida e avaliada a sensibilidade de um ensaio multiplex de 24 polimorfismos (23 SNPs e 1 INDEL) em uma única reação para predição de cor de olhos e cabelos, utilizando uma quantidade diminuta de DNA (63 picogramas). Este sistema foi denominado HIrisplex, sendo o primeiro sistema de predição simultânea de cor dos olhos e cabelos baseados no DNA. Os artigos precursores deste para a criação do conjunto de SNPs preditores de cor dos olhos (Irisplex – Walsh *et al*., 2011) e posterior acréscimo de um conjunto de SNPs para predição também de cor de cabelos (HIrisplex – Walsh *et al*., 2013) foram Liu *et al.* (2009) e Branicki *et al.* (2011), respectivamente. Portanto, o ensaio HIrisPlex utiliza 24 polimorfismos de DNA previamente estabelecidos com um valor preditivo sobre a cor de olhos e cabelos humanos. Os polimorfismos utilizados neste ensaio são: N29insA, rs11547464, rs885479, rs1805008, rs1805005, rs1805006, rs1805007, rs1805009, Y152OCH, rs2228479, rs1110400 – todos do gene *MC1R*; rs12913832 do gene *HERC2*; rs12203592 do gene *IRF4*; rs1042602 e rs1393350 do gene *TYR*; rs4959270 do gene *EXOC2*; rs28777 e rs16891982 do gene *SLC45A2* (*MATP*); rs683 do gene *TYRP1*; rs1800407 do gene *OCA2*; rs2402130 e 12896399 do gene *SLC24A4*; rs12821256 do gene *KITLG*; e o

rs2378249 do gene *ASIP*. Os autores deste artigo disponibilizam uma tabela no *Microsoft Excel*, onde é possível inserir os 24 genótipos de um indivíduo para estes marcadores, e a partir daí, é gerado um *output* indicando qual é a mais provável característica de pigmentação de olhos e cabelos da amostra desconhecida.

Um fato interessante de ser mencionado sobre a predição de características físicas diz respeito à discussão ética que o tema suscita. Uma comissão de Genética Humana do Reino Unido elaborou um relatório que levanta questões éticas sobre o uso da informação genética para predizer as características de uma pessoa através da análise de DNA. Algumas destas considerações foram mencionadas por Tully (2007) e Schneider (2012). O principal argumento em favor da fenotipagem forense pelo DNA é que a pigmentação da pele, olhos e cabelos (por exemplo), por serem características visíveis externamente, não necessitariam de confidencialidade, uma vez que são fenótipos óbvios vistos por qualquer pessoa (Budowle & Van Daal, 2008). Além disso, a fenotipagem forense (que inclui principalmente o uso de SNPs) para desvendar características de pigmentação, sexo, altura, idade, bem como outras características do indivíduo, é útil como um fator investigativo e também como uma poderosa ferramenta de inteligência (Budowle & Van Daal, 2008; e Kayser & Knijff, 2011). Na realidade, a discussão é um pouco mais complexa, pois muitos genes de pigmentação são também preditores para suscetibilidade ao câncer de pele e outras patologias. Além disso, qualquer ligação entre DNA e caracteres fenotípicos aos olhos do especialista em direito e legislação, porém leigo em genética, pode acarretar uma preocupação imediata quanto à eugenia e outros problemas históricos de segregação decorrentes do mau uso de dados biológicos, pelo menos em sociedades onde o estado-de-direito não é plenamente garantido.

Em contrapartida, é bem mais razoável imaginar que uma prova técnico-científica (perfil do suspeito oriundo de um teste de fenotipagem a partir do DNA coletado na cena de um crime, por exemplo) é menos sujeita a erros do que o retrato falado originado pela descrição subjetiva de testemunhas (Spinney, 2008; Kayser & Schneider, 2009; Kayser & Knijff, 2011). Neste aspecto, pode-se antever que menos injustiças serão cometidas e que menos recursos públicos serão despendidos na busca de criminosos quando técnicas de predição de fenótipos visíveis estiverem sendo utilizadas na rotina forense. Algo similar já aconteceu nos EUA visto que nos últimos 38 anos nada menos do que 138 pessoas tiveram suas penas de morte revertidas após terem sua inocência comprovada através de provas

geradas com testes de DNA (http://www.innocenceproject.org/; http://veja.abril.com.br/noticia/internacional/eua-falhas-na-justica-fazem-a-pena-de-morte-perder-forca).

Para finalizar é importante destacar que muitos obstáculos técnicos estão sendo superados para que a predição fenotípica para uso forense seja um fato, e que os aspectos éticos e legais relacionados ao tema devem sempre ser discutidos e avaliados por fóruns especializados e pela sociedade civil.

2. JUSTIFICATIVAS E OBJETIVOS

2.1 Justificativas

Atualmente estamos diante do enorme desafio de conseguir superar as dificuldades na utilização dos complexos e crescentes bancos de dados públicos de informações genéticas. A exploração destas bases de dados (HAPMAP, NCBI, ENSEMBL, UCSC, CEPH) que disponibilizam informações gratuitas sobre sequências e variantes genéticas relacionadas com características normais e patológicas humanas pode trazer enormes benefícios para a comunidade científica, desde que corretamente analisadas. Precisamos deste modo, romper barreiras e montar estratégias e metodologias capazes de responder questões biológicas específicas e solucionar desafios do mundo contemporâneo utilizando estes bancos. Parte dos resultados do presente trabalho se relaciona com a exploração destes dados e sua aplicação para atingir os objetivos propostos.

Além disso, a presente tese também buscou obter dados originais visto que a cor da pele, olhos e cabelos são bastante diferentes em Africanos, Europeus e Nativo-americanos, bem como em populações originárias a partir de distintas dinâmicas de mestiçagem que envolveram indivíduos pertencentes a estes três grupos continentais. Evidências científicas indicam que grupos geográficos diferentes apresentam distintos conjuntos de variantes em genes da rota da pigmentação, como mencionado anteriormente. Deste modo, uma das principais justificativas para o desenvolvimento desta tese é o fato de que ainda são poucos os estudos que têm como foco populações miscigenadas. Diante desta questão, esta tese pretende expandir as análises em populações de estados brasileiros com diferentes históricos de colonização e mestiçagem (predominantemente Bahia e Rio Grande do Sul). Em populações como estas, espera-se que variantes-causais de algumas características ocorrerão mais frequentemente em segmentos cromossômicos herdados a partir de um ou de outro grupo parental. Além disso, tais variantes podem estar em desequilíbrio de ligação com outros alelos funcionais desconhecidos que são mais freqüentes ou mesmo exclusivos no *background* genético de um determinado grupo populacional, mas não em outro. Neste contexto, a utilização de populações miscigenadas pode agregar valor e aumentar o nível de acurácia das informações sobre relação genótipo-fenótipo e o quanto estas podem ser universais ou população-específica.

Os dados descritos aqui possuem relevância médica, evolutiva e forense. Médica porque os estudos de variantes polimórficas que explicam parte da variabilidade genética normal também podem explicar aspectos patológicos de doenças humanas. Na verdade, sabe-se que um aspecto importante para o entendimento da genética de doenças complexas é entender inicialmente como os genes e suas variantes estão relacionados com a varibilidade normal. Isso se aplica muito bem ao estudo da pigmentação humana, uma vez que já se conhece que a suscetibilidade a várias patologias (câncer de pele, por exemplo) estaria associada a alelos associados com a variabilidade normal.

A segunda importante implicação é com relação aos aspectos evolutivos. A diversidade da pigmentação humana tem sido associada a pressões seletivas diversas vivenciadas pelos humanos modernos após sua saída da África. Hoje, com as publicações dos genomas de hominíneos extintos, *Homo neanderthalensis* e o espécime de Denisova (hominínio asiático com taxonomia ainda não definida; Green *et al.*, 2010; Reich *et al*., 2010), é possível explorar a similaridade destes com relação aos humanos modernos no que diz respeito a uma série de traços de interesse. Além disso, é possível também predizer como estes hominíneos seriam no que concerne às suas características físicas.

Finalmente, com relação aos aspectos forenses (a terceira e última implicação de um estudo como este), é esperado que se possa melhorar o entendimento dos fatores envolvidos na relação genótipo-fenótipo da pigmentação humana em populações mestiças, a fim de que esses dados possam auxiliar na predição de características visíveis externamente para fins investigativos policiais, minimizando a possibilidade de falsos acusamentos em processos criminais.

2.2 Objetivos

1) Averiguar se as variantes encontradas estão associadas com traços fenotípicos relacionados com pigmentação de pele, olhos e cabelos na amostra oriunda da nossa população miscigenada;
2) Estimar as freqüências gênicas nos polimorfismos *ADAM17*rs1524668, *AFG3L1*rs4785763, *ASIP*rs6058017, *HERC2*rs1129038, *MC1R*rs1805009, *MC1R*rs1805008, *MC1R*rs1805007, *OCA2*rs1800407, *OCA2*rs1800401, *OCA2*rs1800414, *SLC24A5*rs1426654, *SLC45A2*rs6867641, *SLC45A2*rs26722, *SLC45A2*rs16891982, *TPCN2*rs3750965, *TPCN2*rs3829241, *TYR*rs1042602 e *TYR*rs1126809 em amostras oriundas de dois estados brasileiros (Bahia e Rio grande do Sul), bem como descrever os possíveis efeitos dos alelos analisados;
3) Utilizar banco de dados públicos para explorar as informações pertinentes na área da genética da pigmentação, e tentar estabelecer um método de predição de fenótipos a partir destas informações.

3. RESULTADOS

Os resultados serão apresentados na forma de artigos científicos já publicados, em processo de revisão ou em fase final de preparação para submissão.

3.1. CAPÍTULO I

“Predicting *Homo* pigmentation phenotype through genomic data: from Neanderthal to James Watson**”**

(Artigo Publicado na revista *American Journal of Human Biology*, 24:705-709, 2012)

Short Report

Predicting *Homo* Pigmentation Phenotype Through Genomic Data: From Neanderthal to James Watson

CAIO C. S. CERQUEIRA, VANESSA R. PAIXÃO-CÔRTES, FRANCIS M. B. ZAMBRA, FRANCISCO M. SALZANO, TÁBITA HÜNEMEIER, AND MARIA-CÁTIRA BORTOLINI*
Programa de Pós-Graduação em Genética e Biologia Molecular and Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil

Background: Human pigmentation is regulated by several genes acting at different stages of melanin formation. Functional and association studies have elucidated the role of several of these genes in pigmentation phenotypes. Forensic and evolutionary studies can benefit from this knowledge.

Objectives: To evaluate the reliability of the prediction of pigmentation phenotypes using a large database of genetic markers in individuals with known phenotypes; and from this try to predict the pigmentation phenotype of prehistoric *Homo* specimens and of contemporary individuals whose visible phenotypes are not known.

Methods: We compared predicted and observed phenotypic data through an analysis of 124 single nucleotide polymorphisms in 33 genic and seven intergenic regions of 30 subjects, five of them prehistoric, whose complete nuclear genomes are available in UCSC and PSU UCSC public databases.

Results: For the molecular predicted versus observed phenotypes, the percentage of agreement was as follows: freckles: 91; skin: 64; hair: 44; eyes: 36; total: 59; while the molecular predicted versus probable (no visible observation available; inferences based on ethnic population characteristics) it was, respectively, 83, 60, 42, 67, and 63. The difference between two sets is statistically nonsignificant ($P = 0.75$).

Conclusion: To our knowledge, this is the first article to examine the effect of a large number of genetics markers for phenotype prediction. The approach could be useful for forensic applications, as well as for the determination of possible phenotypes of extinct prehistoric individuals. Am. J. Hum. Biol. 24:705–709, 2012.

BACKGROUND

Extensive studies have identified several genes related with human pigmentation phenotypes (Baxter et al., 2009; Branicki et al., 2011; Nan et al., 2009; Norton et al., 2007; Spichenok et al., 2011; Sulem et al., 2008). These genes include transcription factors, melanogenesis, stabilization, and transport of enzymes during melanin synthesis, membrane and structural proteins, melanosome maintenance, expression of different melanin types, and several kinds of receptors and their ligands (Stinchcombe et al., 2004). Regardless of the amount of studies already published, we are still far from understanding the complex relationship between genotype and human pigmentation.

The recent availability of genotype/phenotype and whole genome databases enables an increase in the use of molecular markers (e.g., SNPs) to estimate externally visible characteristics (Kayser and Schneider, 2009). Today, the amount of information available can allow researchers to reconstruct a "molecular photofitting" from the DNA samples (Bouakaze et al., 2009; Frudakis, 2008; Walsh et al., 2011a). Pigmentation-related genes can be useful in several areas beyond forensic sciences. This approach makes possible estimates of possible phenotypes from early hominids, as well their diversity, rescuing human evolutionary history through previously described phenotype markers (Lalueza-Fox et al., 2007).

Here, we compared predicted and observed phenotypic data through an analysis of 124 single nucleotide polymorphysms in 33 genic and seven intergenic regions of 11 subjects whose personal complete genomes are available in public databases, of five hominids (three *Homo neanderthalensis*, the Denisova and one paleo-Eskimo specimens) plus 14 additional anonymous individuals from the main geographical areas of human extant groups whose DNA was also completely sequenced.

METHODS

Genotype datasets

We selected from the literature 124 single nucleotide polymorphisms associated with pigmentation located in 33 genes and seven intergenic regions (Supporting Information Table 1). Afterward, the UCSC (http://genome.ucsc.edu/cgi-bin/hgGateway - NCBI36/hg18) and PSU UCSC (http://main.genome-browser.bx.psu.edu/cgi-bin/hgGateway - NCBI36/hg18) databases were accessed and the genotype data of 11 individuals (1–11; Tables 1 and 2) who have their complete genome sequence and public photography available were considered. We also recovered, for the same SNPs, genotypes of five extinct hominids plus 14 anonymous individuals (12–30; Tables 1 and 2) from different ethnic groups of Africa, Asia, Europe, New Guinea, Melanesia and America (Supporting Information Tables 2 and 3). For these 19 individuals, we did not have phenotype information available.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Maria-Cátira Bortolini, Programa de Pós-Graduação em Genética e Biologia Molecular and Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: maria.bortolini@ufrgs.br

Received 2 January 2012; Revision received 25 January 2012; Accepted 7 February 2012

DOI 10.1002/ajhb.22263

Published online 12 March 2012 in Wiley Online Library (wileyonlinelibrary.com).

TABLE 1. Number of pigmentation associated alleles in homozygosis for each feature of the subjects

No	Individuals	Fairer skin	Darker skin	Darker brown hair	Lighter brown hair	Brown hair	Not-brown hair	Blond hair	Not-blond hair	Red hair	Not-red hair	Brown eyes	Not-brown eyes	Green eyes	Not-green eyes	Blue eyes	Not-blue eyes	Freckles	Not-freckles	Total homozygotes
1	James Watson (CSHL)	0	0	0	0	2	0	2	0	2	1	0	0	0	0	6	0	2	2	13
2	J. Craig Venter (JCVI)	2	0	0	2	0	1	4	2	10	8	0	4	0	4	9	0	4	15	39
3	Seong-Jin Kim (SJK, GUMS/ KOBIC)	2	0	0	0	0	2	4	1	10	14	2	3	1	4	9	1	6	10	42
4	Misha Angrist (personal genome project)	0	0	0	0	1	0	1	3	8	11	1	2	0	2	1	3	3	16	35
5	George Church (personal genome project)	4	0	0	2	0	2	3	0	10	4	0	4	0	4	7	1	2	10	34
6	Henry Louis Gates Jr. (personal genome project)	1	1	2	0	0	0	1	2	10	7	3	1	2	2	1	1	4	11	29
7	Henry Louis Gates Sr. (personal genome project)	1	2	3	0	0	1	0	2	10	8	6	2	4	3	3	3	9	12	42
8	Rosalynn Gill (personal genome project)	2	0	0	2	0	1	3	1	15	2	0	2	0	2	9	0	6	7	34
9	Marjolein Kriek (Leiden University Medical Centre)	2	0	0	2	0	2	3	2	6	10	0	4	0	2	11	0	2	14	34
10	Gregory Lucier (Life Technologies)	2	0	0	1	0	1	3	1	7	7	0	2	0	3	5	0	0	14	30
11	Stephen Quake (Stanford)	2	0	0	1	0	0	2	1	4	6	0	2	0	1	9	0	4	14	32
12	Neanderthal (Vi33.16)	0	2	1	0	6	0	2	1	12	10	8	2	5	2	2	3	7	10	47
13	Neanderthal (Vi33.25)	0	6	2	0	3	0	1	2	14	9	6	4	1	4	2	4	12	9	48
14	Neanderthal (Vi33.26)	1	3	1	0	6	0	3	2	12	10	7	1	3	2	3	4	8	8	44
15	Denisova	1	13	3	0	10	1	4	10	24	31	19	6	4	8	5	10	18	23	109
16	Sardinian (HGDP00665)	6	6	1	3	4	4	7	7	23	22	6	10	1	8	15	2	2	27	88
17	Papuan (HGDP00542)	7	7	1	4	4	2	9	4	21	21	4	9	2	6	17	5	12	18	84
18	Papuan (HGDP00551)	5	10	2	2	8	3	6	6	19	24	11	6	4	4	10	5	17	18	89
19	Melanesian (HGDP00491)	4	9	2	2	5	2	6	8	29	21	14	7	5	6	8	8	19	18	93
20	San (HGDP01029)	4	8	2	2	6	3	4	9	25	18	13	6	5	5	8	6	14	18	86
21	Yoruba (HGDP00927)	5	7	1	2	5	3	7	6	26	15	9	7	4	4	7	2	13	17	75
22	Mbuti Pygmy (HGDP00456)	1	9	3	1	8	2	2	9	18	21	16	5	6	4	4	7	12	22	89
23	Han (HGDP00778)	5	8	4	0	7	2	1	8	14	19	15	5	6	4	2	7	8	20	76
24	Cambodian (HGDP00711)	3	7	2	1	4	4	5	7	27	16	14	4	4	5	7	8	16	17	89
25	Native American (HGDP00998)	5	3	1	2	5	1	4	4	14	22	5	4	2	2	12	4	14	13	69
26	Mongolian (HGDP01224)	7	5	1	3	6	2	3	6	9	15	11	7	1	6	7	5	6	18	67
27	Han Chinese (YanHuang Project)	2	2	2	0	1	2	2	2	7	8	4	3	3	4	3	1	6	11	35
28	Anonymous Korean individual, AK1 (Genomic Medicine Institute)	1	2	2	0	1	1	1	1	5	8	7	1	5	3	1	3	5	9	31
29	Anonymous Irish Male	3	0	0	2	4	2	3	1	16	5	0	3	0	3	11	0	10	12	53
30	Extinct Paleo-Eskimo (Saqqaq Genome Project)	1	1	1	0	1	1	0	1	10	7	4	0	2	2	0	1	7	5	22

TABLE 2. Observed or probable and predicted phenotypes, according to gene functions

		Freckles		Skin		Hair		Eyes	
Nº	Individuals	Observed or probable phenotype	Molecular predicted phenotype	Observed or probable phenotype	Molecular predicted phenotype	Observed or probable phenotype	Molecular Predicted Phenotype	Observed or probable phenotype	Molecular predicted phenotype
1	James Watson (CSHL)	Yes	Undetermined	Fairer	Undetermined	Unavailable	Brown or Blond	Blue	Blue
2	J. Craig Venter (JCVI)	No	No	Fairer	Fairer	Brown	Blond or Red	Blue	Blue
3	Seong-Jin Kim (SJK, GUMS/ KOBIC)	No	No	Fairer	Fairer	Black	Blond	Black	Blue
4	Misha Angrist (personal genome project)	No	No	Fairer	Undetermined	Brown	Brown[#]	Brown	Undetermined
5	George Church (personal genome project)	No	No	Fairer	Fairer	Red	Red	Green	Blue
6	Henry Louis Gates Jr. (personal genome project)	No	No	Darker	Undetermined	Brown	Red	Brown	Brown
7	Henry Louis Gates Sr. (personal genome project)	No	No	Fairer	Darker[#]	Unavailable	Red	Black	Brown
8	Rosalynn Gill (personal genome project)	No	No[#]	Fairer	Fairer	Red	Red	Green	Blue
9	Marjolein Kriek (Leiden University Medical Centre)	No	No	Fairer	Fairer	Blond	Blond[#]	Blue	Blue
10	Gregory Lucier (Life Technologies)	No	No	Fairer	Fairer	Black	Blond	Black	Blue
11	Stephen Quake (Stanford)	No	No	Fairer	Fairer	Brown	Blond[#]	Green	Blue
12	Neanderthal (Vi33.16)	Unavailable	No	Unavailable	Darker	Unavailable	Darker[#] Brown	Unavailable	Brown
13	Neanderthal (Vi33.25)	Unavailable	Yes	Unavailable	Darker	Unavailable	Red	Unavailable	Brown
14	Neanderthal (Vi33.26)	Unavailable	Undetermined	Unavailable	Darker	Unavailable	Darker[#] Brown	Unavailable	Brown
15	Denisova	Unavailable	No	Unavailable	Darker	Unavailable	Darker Brown	Unavailable	Brown
16	Sardinian (HGDP00665)[*]	Varied	No	Varied	Undetermined	Varied	Red[#]	Varied	Blue
17	Papuan (HGDP00542)[*]	No	No	Darker	Undetermined	Brown/black	Blond	Brown/black	Blue
18	Papuan (HGDP00551)[*]	No	No[#]	Darker	Darker	Brown/black	Brown	Brown/black	Brown or Blue
19	Melanesian (HGDP00491)[*]	No	Yes[#]	Darker	Darker	Brown/black	Red	Brown/black	Brown
20	San (HGDP01029)[*]	No	No	Darker	Darker	Brown/black	Red	Brown/black	Brown
21	Yoruba (HGDP00927)[*]	No	No	Darker	Darker	Brown/black	Red	Brown/black	Blue
22	Mbuti Pygmy (HGDP00456)[*]	No	No	Darker	Darker	Brown/black	Darker Brown	Brown/black	Brown
23	Han (HGDP00778)[*]	No	No	Fairer	Darker	Brown/black	Darker Brown	Brown/black	Brown
24	Cambodian (HGDP00711)[*]	No	No[#]	Unavailable	Darker	Brown/black	Red	Brown/black	Brown
25	Native American (HGDP00998)[*]	No	Yes[#]	Unavailable	Fairer	Brown/black	Lighter[#] Brown	Brown/black	Blue
26	Mongolian (HGDP01224)[*]	No	No	Fairer	Fairer	Brown/black	Lighter Brown	Brown/black	Brown
27	Han Chinese Individual (YanHuang Project)[*]	No	No	Fairer	Undetermined	Brown/black	Undetermined	Brown/black	Blue
28	Anonymous Korean Individual, AK1 (Genomic Medicine Institute)[*]	No	No	Fairer	Darker	Brown/black	Undetermined	Brown/black	Brown
29	Anonymous Irish Male[*]	Varied	No	Varied	Fairer	Varied	Red	Varied	Blue
30	Extinct Paleo-Eskimo (Saqqaq Genome Project)	Unavailable	Yes	Unavailable	Undetermined	Unavailable	Red	Unavailable	Brown

[*]Probable phenotype pigmentation characteristics are based on phenotypes of the majority of individuals from the population to which the individual belongs. Undetermined represents those cases where there is no homozygote or when the number of homozygotes is equal for having or not having the trait, while unavailable represents those cases where phenotype is unclear. Since undetermined cases are related to the method used here, they were included in the calculation of the percentage of correct predictions.
[#]These assignments should be considered with caution since they indicate a difference of only one homozygote between having and not having the trait in question.

General information about the SNPs used (e.g., synonymous or nonsynonymous, intron or exon, base variation, ancestral alelle) was provided by NCBI, entrez SNP (http://www.ncbi.nlm.nih.gov/SNP/) and Ensembl (http://www.ensembl.org/index.html).

Phenotype prediction

We considered 18 phenotypic characteristics (Table 1) in those 30 individuals with public complete nuclear genomic information. One way to predict characteristics of multifactorial inheritance, based on genetic markers, is to consider additive allele effects. Based on this premise, we conducted the following analysis to determine which phenotype is suggested by the number of homozygous genotypes surveyed in the 124 SNPs. This is a conservative approach, since only homozygous effects were used for the prediction, as adopted by Spichenok et al. (2011). Markers in heterozygosis were excluded from the analysis because it is unclear the relationship which one allele has over the other (e.g., dominance, co-dominance, incomplete dominance). The function of each allele for the characterization of possible phenotype was investigated in the articles found in the scientific literature (Supporting Information Table 4). In the polymorphisms which had a functional information associated to an allele, we considered the opposite effect to exist for the other allele of the SNP in question (for example, in a SNP that changes a G to a T, if the G allele was associated with "Blue eyes," we considered that the T allele would lead to "Not-blue eyes").

To check for the consistency of the predicted phenotype information, we compared the real with the predicted phenotype for the 11 subjects from whom we had phenotype information. When anonymous individuals of the different ethnic groups were considered, we assigned their phenotypes considering the average phenotypes of their populations and also compared them with the phenotypes predicted by molecular markers. The two approaches were then compared using the χ^2 test ($\alpha = 0.05$).

RESULTS AND DISCUSSION

The number of homozygous SNPs genotypes significantly associated with the nine dichotomous characteristics of skin, hair and eye considered range from 13 in James Watson (CSHL) to 109 for the Denisova hominid (Table 1). These results were then incorporated to establish the phenotypes of the 30 subjects.

Table 2 lists the predicted versus observed phenotypes for 11 and predicted versus probable phenotypes for the remainder subjects. We verified that, for some characteristics in few situations, there was a difference of only one homozygote for the assignment of having or not having the trait in question (# in Table 2). These cases should be considered with caution, although they are similar to those which occur when a probability of $P = 0.049$ is obtained in a genetic association study.

The percentage of agreement varied in the two kinds of comparisons. For the observed versus molecular predicted phenotypes (subjects 1-11) they were (in percentages) as follows: freckles: 91 (10 agreements/11 comparisons; more details in Table 2 footnotes); skin: 64; hair: 44; eyes: 36; total: 59; while the molecular predicted versus probable (inferences based on ethnic population characteristics; subjects 16–29) it was, respectively, 83, 60, 42, 67, and 63. The difference between the two sets is small and statistically nonsignificant ($P = 0.75$). The level of prediction obtained from population inferences, therefore, did not differ from that obtained when the individual phenotype is known.

Our numbers are similar with some obtained in previous studies. For instance, Walsh et al. (2011b) found high concordance for blue eyes, but lower values for intermediate colors. Our data showed 100% agreement for blue, but less concordance for brown or green eyes. Intermediate colors, like green, apparently have a more complex type of inheritance and it is more difficult to categorize them phenotypically. On the other hand, Spichenok et al. (2011), analyzing just seven SNPs, obtained a very low error rate for both eye and skin colors. Although they used a different methodology from that used here, and also of those used by Walsh et al. (2011b), the small error rate observed in this study is intriguing.

The archaic hominid and Paleo-Eskimo phenotypes cannot be validated because they are extinct, so that only speculations can be made (subjects 12–15, 30; Tables 1 and 2). Lalueza-Fox et al. (2007) found a *MC1R* gene variant (R307G) in two Neanderthals from Italy and Spain, which could determine pale skin and/or red hair as observed in modern humans. This variant was not present in one of the Neanderthals studied here (Vi33.26, the only one in whom this genomic region was available), and was also absent in the Denisova specimen. Our results indicated that they could had darker skin and red or brown hair (Table 2), suggesting a probable variability in these traits among archaic hominids.

To our knowledge, this is the first article to examine the additive effect of a large number of genetic markers in individuals with available complete nuclear genome sequence, to predict their phenotypes. Our methods and results showed a reasonable predictive value (hit average rate of ~60%), although their application still requires additional studies, since there are several genes whose role in pigmentation should be better understood. For instance, epistasis, or markers involved in a pigmentation feature in one population but not in another, as well as other confounding factors, may hamper the interpretation of results.

LITERATURE CITED

Baxter LL, Loftus SK, Pavan WJ. 2009. Networks and pathways in pigmentation, health, and disease. Wiley Interdiscip Rev Syst Biol Med 1:359–371.

Bouakaze C, Keyser C, Crubézy E, Montagnon D, Ludes B. 2009. Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. Int J Legal Med 123:315–25.

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

Frudakis TN. 2008. Molecular photofitting - predicting ancestry and phenotype using DNA. Burlington, MA: Academic Press. 712p.

Kayser M, Schneider PM. 2009. DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. Forensic Sci. Int. Genet 3:154–161.

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

Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, Hu FB, Thomas G, Hoover RN, Chanock S, Hunter DJ, Han J. 2009. Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol 129:2250–2257.

Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD. 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol 24:710–722.

Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, Caragine T, Prinz M, Wurmbach E. 2011. Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Sci Int Genet 5:472–478.

Stinchcombe J, Bossi G, Griffiths CM. 2004. Linking albinism and immunity: the secrets of secretory lysosomes. Science 305:55–59.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G, Pálsson S, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Vermeulen SH, Goldstein AM, Tucker MA, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K. 2008. Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40:835–837.

Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, Kayser M, Ballantyne KN. 2011a. Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5:464–471.

Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. 2011b. IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5:170–180.

Supporting Information Table 1. Characteristics of the SNPs studied in 33 genes and 7 intergenic regions

N°	Region	Chr.	Position	SNPs ID	Mutation	Type	Ancestral Allele
1	***SLC45A2 (MATP)***	5p13.3	chr5:33,999,627-33,999,627	rs26722	Glu272Lys - G>A	NS	G
	SLC45A2 (MATP)	5p13.3	chr5:34,021,614-34,021,614	rs6867641	C>T	nearGene-5	C
2	***IRF4***	6p25.3	chr6:341,321-341,321	rs12203592	C>T	Intron	C
3	***TPCN2***	11q13.2	chr11:68,611,939-68,611,939	rs3829241	Gly734Glu - G>A	NS	G
	TPCN2	11q13.2	chr11:68,596,736-68,596,736	rs3750965	Lys376Arg - A>G	NS	A
	TPCN2	11q13.2	chr11:68,602,975-68,602,975	rs35264875	Met484Leu - A>T	NS	A
	TPCN2	11q13.2	chr11:68,585,505-68,585,505	rs896978	C>T	Intron	C
4	***TYR***	11q14.3	chr11:88,551,344-88,551,344	rs1042602	Ser192Tyr - C>A	NS	C
	TYR	11q14.3	chr11:88,657,609-88,657,609	rs1126809	Arg402Gln - G>A	NS	G
	TYR	11q14.3	chr11:88,650,694-88,650,694	rs1393350	G>A	Intron	G
5	***SLC24A4***	14q32.12	chr14:91,870,956-91,870,956	rs2402130	G>A	nd	G
6	***OCA2***	15q12	chr15:25,685,246-25,685,246	rs1498519	C>A	Intron	C
	OCA2	15q13.1	chr15:25,870,632-25,870,632	rs1800414	His615Arg - A>G	NS	A
	OCA2	15q13.1	chr15:25,903,913-25,903,913	rs1800407	Arg419Gln - G>A	NS	G
	OCA2	15q13.1	chr15:25,933,648-25,933,648	rs1800401	Arg305Trp - C>T	NS	C
	OCA2	15q13.1	chr15:26,017,833-26,017,833	rs7495174	G>A	Intron	G
	OCA2	15q13.1	chr15:26,012,308-26,012,308	rs4778241	A>C	Intron	A
	OCA2	15q13.1	chr15:26,009,415-26,009,415	rs4778138	G>A	Intron	G
	OCA2	15q13.1	chr15:25,830,854-25,830,854	rs1584407	C>A	Intron	C
	OCA2	15q13.1	chr15:25,855,576-25,855,576	rs2703952	A>C	Intron	A
	OCA2	15q13.1	chr15:25,858,633-25,858,633	rs2594935	A>G	Intron	A
	OCA2	15q13.1	chr15:25,873,448-25,873,448	rs728405	G>T	Intron	G
	OCA2	15q13.1	chr15:25,890,452-25,890,452	rs1448488	A>G	Intron	A
	OCA2	15q13.1	chr15:25,894,733-25,894,733	rs4778220	A>G	Intron	A
	OCA2	15q13.1	chr15:25,962,343-25,962,343	rs7170869	G>A	Intron	G
	OCA2	15q13.1	chr15:25,861,367-25,861,367	rs1545397	A>T	Intron	A

Supporting Information Table 1. Cont

N°	Region	Chr.	Position	SNPs ID	Mutation	Type	Ancestral Allele
7	*HERC2*	15q13.1	chr15:26,030,454-26,030,454	rs1129038	G>A	3'UTR	G
	HERC2	15q13.1	chr15:26,039,213-26,039,213	rs12913832	A>G	Intron	A
	HERC2	15q13.1	chr15:26,203,777-26,203,777	rs1667394	G>A	Intron	G
	HERC2	15q13.1	chr15:26,189,679-26,189,679	rs8039195	C>T	Intron	C
	HERC2	15q13.1	chr15:26,039,328-26,039,328	rs7183877	C>A	Intron	C
	HERC2	15q13.1	chr15:26,208,861-26,208,861	rs1635168	T>G	Intron	T
	HERC2	15q13.1	chr15:26,162,483-26,162,483	rs8028689	C>T	Intron	C
	HERC2	15q13.1	chr15:26,199,823-26,199,823	rs16950987	A>G	Intron	A
	HERC2	15q13.1	chr15:26,186,959-26,186,959	rs916977	A>G	Intron	A
	HERC2	15q13.1	chr15:26,037,654-26,037,654	rs7494942	A>G	Intron	A
	HERC2	15q13.1	chr15:26,047,607-26,047,607	rs3935591	A>G	Intron	A
	HERC2	15q13.1	chr15:26,101,581-26,101,581	rs7170852	T>A	Intron	T
	HERC2	15q13.1	chr15:26,126,810-26,126,810	rs2238289	C>T	Intron	C
	HERC2	15q13.1	chr15:26,167,797-26,167,797	rs2240203	G>A	Intron	G
	HERC2	15q13.1	chr15:26,167,627-26,167,627	rs2240204	C>T	Intron	C
	HERC2	15q13.1	chr15:26,194,101-26,194,101	rs16950979	A>G	Intron	A
8	*SLC24A5 (NCKX5)*	15q21.1	chr15:46,213,776-46,213,776	rs1426654	Ala111Thr - G>A	NS	G
9	*MC1R*	16q24.3	chr16:88,513,618-88,513,618	rs1805007	Arg151Cys - C>T	NS	C
	MC1R	16q24.3	chr16:88,513,645-88,513,645	rs1805008	Arg160Trp - C>T	NS	C
	MC1R	16q24.3	chr16:88,509,859-88,509,859	rs3212346	A>G	nearGene-5	A
	MC1R	16q24.3	chr16:88,513,655-88,513,655	rs885479	Arg163Gln - G>A	NS	G
10	*DPEP1*	16q24.3	chr16:88,219,799-88,219,799	rs164741	T>C	Intron	T
11	*C16orf55*	16q24.3	chr16:88,253,985-88,253,985	rs7188458	G>A	Intron	G
	C16orf55	16q24.3	chr16:88,258,328-88,258,328	rs459920	T>C	Intron	T

Supporting Information Table 1. Cont

N°	Region	Chr.	Position	SNPs ID	Mutation	Type	Ancestral Allele
12	*ZNF276*	16q24.3	chr16:88,322,986-88,322,986	rs7204478	T>C	Intron	T
	ZNF276	16q24.3	chr16:88,317,399-88,317,399	rs6500437	Trp188Arg - T>C	NS	T
	ZNF276	16q24.3	chr16:88,332,762-88,332,762	rs1800359	C>T	3'UTR	C
13	*ZNF778*	16q24.3	chr16:87,821,940-87,821,940	rs9921361	Gln553His - G>T	NS	G
14	*PRDM7*	16q24.3	chr16:88,657,637-88,657,637	rs2078478	C>T	nearGene-5	nd
	PRDM7	16q24.3	chr16:88,668,978-88,668,978	rs7196459	G>T	Intron	G
15	*ACSF3*	16q24.3	chr16:87,733,984-87,733,984	rs12599126	C>T	Intron	C
16	*ANKRD11*	16q24.3	chr16:87,913,062-87,913,062	rs2353033	C>T	Intron	C
	ANKRD11	16q24.3	chr16:87,871,978-87,871,978	rs1466540	C>T	Intron	C
	ANKRD11	16q24.3	chr16:87,880,179-87,880,179	rs2353028	G>A	Intron	G
	ANKRD11	16q24.3	chr16:87,882,779-87,882,779	rs2306633	A>G	Intron	A
	ANKRD11	16q24.3	chr16:87,901,208-87,901,208	rs3096304	G>A	Intron	G
	ANKRD11	16q24.3	chr16:87,914,309-87,914,309	rs889574	C>T	Intron	C
	ANKRD11	16q24.3	chr16:88,044,113-88,044,113	rs2965946	C>T	Intron	C
17	*SPG7*	16q24.3	chr16:88,131,087-88,131,087	rs382745	C>T	3'UTR	C
18	*CPNE7*	16q24.3	chr16:88,171,502-88,171,502	rs455527	Phe77Leu - T>C	NS	T
	CPNE7	16q24.3	chr16:88,176,081-88,176,081	rs352935	G>A	Intron	G
	CPNE7	16q24.3	chr16:88,183,752-88,183,752	rs464349	T>C	Intron	T
19	*CHMP1A*	16q24.3	chr16:88,240,390-88,240,390	rs460879	C>T	Intron	C
20	*CDK10*	16q24.3	chr16:88,283,404-88,283,404	rs258322	C>T	Intron	C
	CDK10	16q24.3	chr16:88,281,756-88,281,756	rs258324	C>A	Intron	C
	CDK10	16q24.3	chr16:88,279,695-88,279,695	rs3751700	G>A	nearGene-5	G
	CDK10	16q24.3	chr16:88,289,911-88,289,911	rs1946482	T>C	nearGene-3	T
21	*SPATA2L*	16q24.3	chr16:88,292,050-88,292,050	rs3751695	Phe156Phe - C>T	S	C

Supporting Information Table 1. Cont

N°	Region	Chr.	Position	SNPs ID	Mutation	Type	Ancestral Allele
22	*FANCA*	16q24.3	chr16:88,363,824-88,363,824	rs7195066	Asp809Gly - A>G	NS	A
	FANCA	16q24.3	chr16:88,342,308-88,342,308	rs8058895	T>C	Intron	T
	FANCA	16q24.3	chr16:88,342,319-88,342,319	rs2011877	A>C>T	Intron	A
	FANCA	16q24.3	chr16:88,376,981-88,376,981	rs2239359	Ser501Gly - A>G	NS	A
	FANCA	16q24.3	chr16:88,378,534-88,378,534	rs16966142	C>T	Intron	C
	FANCA	16q24.3	chr16:88,397,262-88,397,262	rs1800286	G>A	Intron	G
	FANCA	16q24.3	chr16:88,403,211-88,403,211	rs11861084	C>A	Intron	C
23	*SPIRE2*	16q24.3	chr16:88,447,526-88,447,526	rs8060934	C>T	Intron	C
	SPIRE2	16q24.3	chr16:88,462,387-88,462,387	rs3803688	T>C	Intron	T
24	*TCF25*	16q24.3	chr16:88,499,917-88,499,917	rs2270460	G>T	Intron	G
25	*CENPBD1*	16q24.3	chr16:88,565,329-88,565,329	rs4785755	Thr168Ile - C>T	NS	C
26	*DBNDD1*	16q24.3	chr16:88,607,035-88,607,035	rs8059973	G>A	Intron	G
	DBNDD1	16q24.3	chr16:88,612,062-88,612,062	rs11648785	C>T	Intron	C
27	*GAS8*	16q24.3	chr16:88,615,938-88,615,938	rs2241039	T>C	nearGene-5	T
	GAS8	16q24.3	chr16:88,632,834-88,632,834	rs3785181	G>A	Intron	G
	GAS8	16q24.3	chr16:88,638,451-88,638,451	rs1048149	C>T	3'UTR	C
28	*AFG3L1*	16q24.3	chr16:88,594,437-88,594,437	rs4785763	A>C	Pseudogene	A
	AFG3L1	16q24.3	chr16:88,571,529-88,571,529	rs4408545	C>T	Pseudogene	C
29	*DYNLRB1*	20q11.22	chr20:32,592,825-32,592,825	rs2281695	T>C	nearGene-3	T
30	*PIGU*	20q11.22	chr20:32,650,141-32,650,141	rs2378199	T>C	Intron	T
	PIGU	20q11.22	chr20:32,681,751-32,681,751	rs2378249	A>G	Intron	A
31	*NCOA6*	20q11.22	chr20:32,815,525-32,815,525	rs6060034	C>T	Intron	C
	NCOA6	20q11.22	chr20:32,828,245-32,828,245	rs6060043	T>C	Intron	T
32	*EIF6*	20q11.22	chr20:33,331,111-33,331,111	rs619865	G>A	Intron	G
33	*ASIP*	20q11.22	chr20:32,320,659-32,320,659	rs6058017	G>A	3' UTR	G

Supporting Information Table 1. Cont

N°	Region	Chr.	Position	SNPs ID	Mutation	Type	Ancestral Allele
1	Intergenic regions	6p25.3	chr6:379,364-379,364	rs9328192	A>G	-	A
		6p25.3	chr6:394,358-394,358	rs9405681	C>T	-	C
		6p25.3	chr6:402,748-402,748	rs4959270	C>A	-	C
		6p25.3	chr6:362,727-362,727	rs9378805	A>C	-	A
		6p25.3	chr6:411,033-411,033	rs1540771	G>A	-	G
2	Intergenic regions	11q13.2	chr11:68,690,473-68,690,473	rs1011176	G>A	-	G
		11q13.2	chr11:68,623,490-68,623,490	rs2305498	C>T	-	C
3	Intergenic regions	12q21.33	chr12:87,852,466-87,852,466	rs12821256	T>C	-	T
4	Intergenic regions	14q32.12	chr14:91,828,198-91,828,198	rs8016079	G>A	-	G
		14q32.12	chr14:91,834,272-91,834,272	rs4904864	G>A	-	G
		14q32.12	chr14:91,850,754-91,850,754	rs4904868	T>C	-	T
		14q32.12	chr14:91,843,416-91,843,416	rs12896399	G>T	-	G
5	Intergenic regions	15q11.2	nd	rs438702	G>A	-	nd
6	Intergenic regions	16q24.3	chr16:87,580,066-87,580,066	rs9932354	C>A	-	C
		16q24.3	chr16:87,584,526-87,584,526	rs11076747	G>A	-	G
		16q24.3	chr16:87,855,978-87,855,978	rs4785648	A>G	-	A
		16q24.3	chr16:88,098,136-88,098,136	rs4347628	C>T	-	C
		16q24.3	chr16:88,268,997-88,268,997	rs12443954	G>A	-	G
		16q24.3	chr16:88,596,560-88,596,560	rs9936896	C>T	-	C
		16q24.3	chr16:88,640,608-88,640,608	rs4785612	C>A	-	C
		16q24.3	chr16:88,586,247-88,586,247	rs7201721	A>G	-	A
		16q24.3	chr16:88,578,190-88,578,190	rs4238833	G>T	-	G
7	Intergenic regions	20q11.22	chr20:32,248,873-32,248,873	rs6119471	G>C	-	G
		20q11.22	chr20:32,202,273-32,202,273	rs1015362	A>G	-	A
		20q11.22	chr20:32,193,105-32,193,105	rs4911414	G>T	-	G

- = non-applicable; A= Adenine; ACSF3 = acyl-CoA synthetase family member 3; AFG3L1 = ATPase family gene 3 like 1 (Yeast); Ala = alanine; ANKRD11 = ankyrin repeat domain 11; Arg = arginine; ASIP = agouti signaling protein; Asp = aspartic acid; C= Cytosine; C16orf55 = chromosome 16 open reading frame 55; CDK10 = cyclin-dependent kinase 10; CENPBD1 = CENPB DNA-binding domains containing 1; CHMP1A = chromatin modifying protein 1A; CPNE7 = copine VII; Cys = cysteine; DBNDD1 = dysbindin (dystrobrevin binding protein 1) domain containing 1; DPEP1 = dipeptidase 1; DYNLRB1 = dynein, light chain, roadblock-type 1; EIF6 = eukaryotic translation initiation factor 6; FANCA = Fanconi anemia,

complementation group A; G= Guanine; GAS8 = growth arrest-specific 8; Gln = glutamine; Glu = glutamic acid; Gly = glycine; HERC2 = hect domain and RLD 2; His = histidine; Ile = isoleucine; IRF4 = interferon regulatory factor 4; Leu = leucine; Lys = lysine; MATP = membrane associated transporter protein; MC1R = melanocortin 1 receptor; Met = methionine; NCKX5= Na/Ca/K exchanger 5; NCOA6 = nuclear receptor coactivator 6; nd = non-determined; NS = non-synonymous; OCA2 = oculocutaneous albinism II; Phe = phenylalanine; PIGU = phosphatidylinositol glycan anchor biosynthesis, class U; PRDM7 = PR domain containing 7; S = synonymous; Ser = serine; SLC24A4 = solute carrier family 24, member 4; SLC24A5 = solute carrier family 24, member 5; SLC45A2 = solute carrier family 45 member 2; SPATA2L = spermatogenesis associated 2-like; SPG7 = spastic paraplegia 7; SPIRE2 = spire homolog 2 (Drosophila); T= Timine; TCF25 = transcription factor 25; Thr = threonine; TPCN2 = two pore segment channel 2; Trp = tryptophan; Tyr = tyrosine (amino acid); TYR = tyrosinase; UTR = untranslated region; ZNF276 = zinc finger protein 276; ZNF778 = zinc finger protein 778.

Supporting Information Table 2. Genotypes of individuals 1 to 11 searched for pigmentation SNPs in the UCSC and PSU/UCSC databanks

Region	SNPs ID	Subjects										
		1	2	3	4	5	6	7	8	9	10	11
SLC45A2 (MATP)	rs26722	nd	nd	G/A	G/A	nd	nd	nd	nd	nd	nd	nd
SLC45A2 (MATP)	rs6867641	nd	nd	C	nd	C/T	C	nd	nd	C/T	nd	C
IRF4	rs12203592	C/T	nd	nd	nd	C/T	C/T	T	nd	nd	nd	nd
TPCN2	rs3829241	G/A	nd	G/A	G/A	A	G/A	G/A	G/A	nd	A	G/A
TPCN2	rs3750965	G	nd	nd	nd	nd	A/G	A/G	A/G	nd	nd	A/G
TPCN2	rs35264875	nd	nd	nd	nd	nd	nd	nd	nd	A/T	nd	nd
TPCN2	rs896978	nd	nd	nd	C/T	nd	C/T	C/T	nd	nd	nd	C/T
TYR	rs1042602	C/A	nd	nd	nd	A	nd	C/A	nd	nd	nd	nd
TYR	rs1126809	nd	G/A	nd	G/A	nd	nd	nd	nd	A	nd	nd
TYR	rs1393350	nd	G/A	nd	G/A	nd	nd	nd	nd	nd	G/A	A
SLC24A4	rs2402130	G/A	A	A	G/A	A	G/A	G/A	A	nd	A	G/A
OCA2	rs1498519	nd	C/A	nd	nd	nd	nd	nd	nd	A	nd	C/A
OCA2	rs1800414	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
OCA2	rs1800407	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
OCA2	rs1800401	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
OCA2	rs7495174	nd	nd	nd	nd	nd	G/A	G	nd	nd	nd	nd
OCA2	rs4778241	A/C	C	A/C	A/C	C	A/C	nd	C	C	C	A/C
OCA2	rs4778138	nd	nd	G/A	nd	nd	G	G	nd	nd	G/A	G/A
OCA2	rs1584407	nd	nd	A	A	A	C/A	C/A	nd	nd	C/A	C/A
OCA2	rs2703952	nd	nd	nd	nd	nd	nd	A/C	nd	nd	nd	A/C
OCA2	rs2594935	nd	nd	nd	A	nd	A/G	A/G	A/G	nd	nd	A/G
OCA2	rs728405	nd	T	nd	G/T	nd	G/T	T	G/T	nd	nd	nd
OCA2	rs1448488	A/G	A	A	nd	A	A/G	A	A/G	nd	A	A/G
OCA2	rs4778220	nd	A	A	A	A	A	A	A/G	A	nd	A/G
OCA2	rs7170869	nd	nd	G	G/A	nd	G	G	nd	nd	nd	G/A
OCA2	rs1545397	nd	nd	nd	A/T	nd	nd	nd	A/T	nd	nd	nd

Supporting Information Table 2. Cont

Region	SNPs ID	Subjects										
		1	2	3	4	5	6	7	8	9	10	11
HERC2	rs1129038	A	A	nd	nd	A	nd	nd	A	A	A	A
HERC2	rs12913832	A/G	G	nd	nd	G	nd	nd	G	G	nd	G
HERC2	rs1667394	A	A	A	nd	nd	nd	nd	A	A	G/A	A
HERC2	rs8039195	T	T	T	nd	T	C/T	nd	T	T	T	T
HERC2	rs7183877	nd	nd	nd	A	nd	nd	nd	nd	nd	nd	nd
HERC2	rs1635168	T/G	G	G	G	G	T/G	nd	G	G	G	G
HERC2	rs8028689	nd	nd	nd	nd	nd	C/T	C	nd	nd	nd	nd
HERC2	rs16950987	nd	nd	nd	nd	nd	A/G	A	nd	nd	nd	nd
HERC2	rs916977	G	nd	G	nd	G	nd	nd	G	G	A/G	G
HERC2	rs7494942	A/G	G	G	nd	G	nd	nd	G	G	A/G	G
HERC2	rs3935591	G	G	G	nd	G	A/G	nd	G	G	G	G
HERC2	rs7170852	A	A	A	nd	nd	T/A	nd	A	A	A	A
HERC2	rs2238289	nd	nd	nd	C	nd	C/T	C	nd	nd	nd	nd
HERC2	rs2240203	nd	nd	nd	nd	nd	G/A	G	nd	nd	nd	nd
HERC2	rs2240204	nd	nd	nd	nd	nd	C/T	C/T	nd	nd	nd	nd
HERC2	rs16950979	nd	nd	nd	nd	nd	A/G	A/G	nd	nd	nd	nd
SLC24A5 (NCKX5)	rs1426654	nd	nd	G/A	nd	nd	G	G	nd	nd	nd	nd
MC1R	rs1805007	nd	C/T	nd	nd	C/T	nd	nd	nd	nd	nd	nd
MC1R	rs1805008	nd	nd	nd	nd	nd	nd	nd	C/T	nd	nd	nd
MC1R	rs3212346	A/G	nd	nd	A/G	nd	A	A	A/G	nd	nd	nd
MC1R	rs885479	nd	nd	A	nd	nd	nd	nd	nd	nd	nd	nd
DPEP1	rs164741	T/C	T/C	nd	C	T/C	T/C	T/C	T/C	C	C	C
C16orf55	rs7188458	G/A	G/A	nd	G/A	G/A	G/A	A	A	nd	G/A	G/A
C16orf55	rs459920	nd	T/C	nd	T/C	T/C	nd	nd	nd	C	T/C	T/C
ZNF276	rs7204478	T/C	T	T	nd	nd	T	T	T/C	nd	T/C	nd
ZNF276	rs6500437	nd	T/C	C	nd	T/C	C	C	nd	nd	T/C	nd
ZNF276	rs1800359	nd	C	C	C/T	C/T	C	C	C	nd	C/T	nd
ZNF778	rs9921361	nd	nd	T	T	T	T	G/T	T	T	T	T

Supporting Information Table 2. Cont

Region	SNPs ID	Subjects										
		1	2	3	4	5	6	7	8	9	10	11
PRDM7	rs2078478	C/T	nd	T	nd	nd	nd	nd	nd	nd	nd	nd
PRDM7	rs7196459	G/T	G	G	G	G	G	G	G/T	G	G	G/T
ACSF3	rs12599126	nd	C/T	C/T	nd	nd	C/T	C	nd	nd	nd	nd
ANKRD11	rs2353033	C/T	T	C/T	C/T	C/T	nd	C/T	nd	C/T	T	T
ANKRD11	rs1466540	nd	C/T	nd	T	T	C/T	C/T	T	nd	T	C/T
ANKRD11	rs2353028	G/A	A	nd	A	A	G/A	G/A	A	nd	nd	G/A
ANKRD11	rs2306633	A/G	G	nd	G	G	A/G	A/G	G	A/G	G	A/G
ANKRD11	rs3096304	nd	A	nd	A	nd	nd	nd	A	A	A	G/A
ANKRD11	rs889574	C	C	C/T	C/T	C/T	C	C	C/T	C/T	C	C
ANKRD11	rs2965946	C	nd	C	C/T	C/T	C/T	C	C/T	C	C	C
SPG7	rs382745	nd	nd	C/T	nd	nd	nd	C/T	nd	nd	nd	C/T
CPNE7	rs455527	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
CPNE7	rs352935	nd	G/A	G/A	G/A	G/A	nd	nd	nd	nd	A	A
CPNE7	rs464349	nd	T/C	C	T/C	C	nd	nd	C	C	nd	T/C
CHMP1A	rs460879	nd	C/T	T	T	C/T	T	C/T	nd	nd	nd	C/T
CDK10	rs258322	C/T	C/T	nd	C	C/T	C/T	C	C	C	C/T	C
CDK10	rs258324	nd	C	C	C	C	C	C	C	C	C	C
CDK10	rs3751700	G/A	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
CDK10	rs1946482	C	nd	nd	T/C	T/C	nd	nd	nd	T/C	nd	nd
SPATA2L	rs3751695	C/T	nd	nd	nd	nd	nd	nd	nd	T	nd	nd
FANCA	rs7195066	A/G	A/G	A	nd	nd	A	A	A/G	nd	nd	nd
FANCA	rs8058895	nd	nd	nd	nd	nd	nd	nd	T/C	nd	nd	nd
FANCA	rs2011877	A/C	A/C/T	C	nd	A/C	C	C	nd	nd	A/C	nd
FANCA	rs2239359	A/G	nd	A/G	A/G	nd	A/G	A/G	A	nd	nd	A/G
FANCA	rs16966142	C/T	nd	C/T	nd	nd	nd	nd	nd	nd	nd	nd
FANCA	rs1800286	G/A	G	G	G/A	G/A	G	G	G	nd	G/A	G/A
FANCA	rs11861084	C	C	C	C/A	C/A	C	C	C	nd	C/A	C/A

Supporting Information Table 2. Cont

Region	SNPs ID	Subjects										
		1	2	3	4	5	6	7	8	9	10	11
SPIRE2	rs8060934	nd	C/T	T	nd	C/T	C/T	nd	C/T	nd	nd	T
SPIRE2	rs3803688	T/C	nd	C	nd	nd	T/C	nd	nd	nd	nd	nd
TCF25	rs2270460	G/T	nd	nd	nd	nd	G/T	nd	nd	nd	nd	nd
CENPBD1	rs4785755	C/T	T	T	T	T	C/T	C/T	C/T	T	T	C/T
DBNDD1	rs8059973	nd	G	nd	G	G	G	G	G	nd	nd	G
DBNDD1	rs11648785	nd	C/T	nd	nd	nd	nd	nd	nd	nd	nd	T
GAS8	rs2241039	T/C	T	T	nd	nd	nd	nd	nd	nd	nd	T/C
GAS8	rs3785181	nd	nd	A	nd	nd	nd	nd	nd	nd	nd	nd
GAS8	rs1048149	C/T	C/T	nd	nd	nd	C/T	C/T	C/T	nd	C/T	nd
AFG3L1	rs4785763	A/C	C	A/C	C	nd	A/C	A/C	A/C	C	A/C	A/C
AFG3L1	rs4408545	nd	T	nd	T	C/T	nd	nd	C/T	T	C/T	C/T
DYNLRB1	rs2281695	T/C	C	T/C	C	C	C	C	C	nd	C	C
PIGU	rs2378199	T/C	C	T/C	C	C	C	C	C	C	C	C
PIGU	rs2378249	nd	A	A/G	A	A	A	A	A	nd	A	A
NCOA6	rs6060034	nd	C	C/T	C	C	C	C	nd	C	C	C
NCOA6	rs6060043	T/C	T	T/C	T	T	T	T	T	T	T	T
EIF6	rs619865	G/A	G	G	G	G	G	G	G	G/A	G	G
ASIP	rs6058017	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
Intergenic regions	rs9328192	nd	nd	A/G	A/G	A/G	A/G	G	A/G	nd	nd	G
	rs9405681	nd	C/T	C/T	C	C/T	C/T	C	C/T	C/T	C/T	C
	rs4959270	C/A	nd	C/A	A	C/A	C/A	A	C/A	nd	nd	nd
	rs9378805	A/C	nd	A/C	nd	C	A/C	C	nd	A/C	A/C	C
	rs1540771	G/A	nd	G/A	A	G/A	nd	G/A	G/A	nd	nd	nd
	rs1011176	G	nd	G/A	G/A	nd	G/A	G/A	G/A	nd	nd	nd
	rs2305498	nd	nd	nd	nd	nd	nd	nd	nd	C/T	nd	nd
	rs12821256	nd	C	nd	nd	nd	nd	nd	nd	T/C	nd	nd

Supporting Information Table 2. Cont

Region	SNPs ID	Subjects										
		1	2	3	4	5	6	7	8	9	10	11
Intergenic regions	rs8016079	G/A	nd	nd	nd	nd	nd	nd	nd	nd	nd	G/A
	rs4904864	nd	G/A	G	G	G/A	G	G/A	G/A	nd	G/A	G/A
	rs4904868	T/C	T/C	C	T/C	T/C	T/C	C	T/C	C	T/C	nd
	rs12896399	G/T	G/T	T	G/T	G/T	G/T	G/T	G/T	G/T	nd	nd
	rs438702	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
	rs9932354	nd	nd	nd	nd	nd	nd	nd	A	nd	nd	nd
	rs11076747	nd	nd	nd	nd	A	nd	nd	A	nd	nd	G/A
	rs4785648	A/G	G	A/G	G	G	G	G	G	G	G	A/G
	rs4347628	C/T	nd	C/T	nd	C/T	nd	nd	C/T	C/T	C/T	C/T
	rs12443954	nd	nd	G	nd	nd	nd	G/A	G/A	nd	nd	nd
	rs9936896	C	C/T	nd	nd	nd	nd	nd	C/T	nd	nd	nd
	rs4785612	C/A	C/A	A	A	A	C/A	C/A	C/A	A	C/A	nd
	rs7201721	A/G	A/G	A	A	A	A	A	A/G	nd	A	A/G
	rs4238833	G/T	T	T	T	G/T	G/T	G/T	G/T	T	G/T	G/T
	rs6119471	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
	rs1015362	nd	nd	nd	nd	A/G	A/G	nd	nd	nd	nd	nd
	rs4911414	nd	G	G	G	G/T	G	G	G	G	G	G

1= James Watson (CSHL); 2= J. Craig Venter (JCVI); 3= Seong-Jin Kim (SJK, GUMS/KOBIC); 4= Misha Angrist (Personal Genome Project); 5= George Church (Personal Genome Project); 6= Henry Louis Gates Jr. (Personal Genome Project); 7= Henry Louis Gates Sr. (Personal Genome Project); 8= Rosalynn Gill (Personal Genome Project); 9= Marjolein Kriek (Leiden University Medical Centre); 10= Gregory Lucier (Life Technologies); 11= Stephen Quake (Stanford); A= Adenine; ACSF3 = acyl-CoA synthetase family member 3; AFG3L1 = ATPase family gene 3 like 1 (Yeast); ANKRD11 = ankyrin repeat domain 11; ASIP = agouti signaling protein; C= Cytosine; C16orf55 = chromosome 16 open reading frame 55; CDK10 = cyclin-dependent kinase 10; CENPBD1 = CENPB DNA-binding domains containing 1; CHMP1A = chromatin modifying protein 1A; CPNE7 = copine VII; DBNDD1 = dysbindin (dystrobrevin binding protein 1) domain containing 1; DPEP1 = dipeptidase 1; DYNLRB1 = dynein, light chain, roadblock-type 1; EIF6 = eukaryotic translation initiation factor 6; FANCA = Fanconi anemia, complementation group A; G= Guanine; GAS8 = growth arrest-specific 8; HERC2 = hect domain and RLD 2; IRF4 = interferon regulatory factor 4; MATP = membrane associated transporter protein; MC1R = melanocortin 1 receptor; NCKX5= Na/Ca/K exchanger 5; NCOA6 = nuclear receptor coactivator 6; nd = non-determined; OCA2 = oculocutaneous albinism II; PIGU = phosphatidylinositol glycan anchor biosynthesis, class U; PRDM7 = PR domain containing 7; SLC24A4 = solute carrier family 24, member 4; SLC24A5 = solute carrier family 24, member 5; SLC45A2 = solute carrier family 45 member 2; SPATA2L = spermatogenesis associated 2-like; SPG7 = spastic paraplegia 7; SPIRE2 = spire homolog 2 (Drosophila); T= Timine; TCF25 = transcription factor 25; TPCN2 = two pore segment channel 2; TYR = tyrosinase; ZNF276 = zinc finger protein 276; ZNF778 = zinc finger protein 778.

Supporting Information Table 3. Genotypes of individuals 12 to 30 searched for pigmentation SNPs in the UCSC and PSU/UCSC databanks

Region	SNPs ID	Subjects																		
		12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
SLC45A2 (MATP)	rs26722	nd	nd	nd	nd	G	G	G	G	G	G	G	G/A	G	G	G	nd	nd	nd	nd
SLC45A2 (MATP)	rs6867641	nd	nd	C	C	C/T	T	T	C/G	C	C	C/T	C	C	C	C	C	C/T	C	nd
IRF4	rs12203592	nd	C	nd	C	C	C/A	C	C	C/A	C	C	C	C	nd	C	nd	nd	C/T	C/T
TPCN2	rs3829241	nd	nd	nd	G	G/A	G	G	G	G	G/A	G	G	G	nd	G	G/A	A	nd	G/A
TPCN2	rs3750965	A	A	A	A	G	A	G	A	A	A	A	G	A	nd	A/G	nd	nd	G	nd
TPCN2	rs35264875	A	A	nd	A	A	A/T/G	A	A	A	A	A	A	A	nd	A	nd	nd	nd	nd
TPCN2	rs896978	nd	nd	nd	C	C/T	C/T	T	C/T	C	C	C	C/T	C	C	T	nd	nd	T	nd
TYR	rs1042602	nd	C	C	C	A	C	C	C	C	C	C	C	C	C	A	nd	nd	nd	nd
TYR	rs1126809	G	nd	G	G	G	G	G	G	G	G/A	G	G	A	G	G	nd	nd	A	nd
TYR	rs1393350	G	nd	G	G	nd	G	G	G	G	G	G	G	G	G	G	nd	nd	A	nd
SLC24A4	rs2402130	G	nd	G	G	A	A	A	A	G/T	G	G	G/A	A	nd	G/A	A	A	A	A
OCA2	rs1498519	nd	C	nd	C	C	C	C	C	C	C	C	C	C	C	C	nd	nd	nd	nd
OCA2	rs1800414	nd	A	nd	A	A	A	A	A	A	A/C	A	G	A/G	nd	A/G	A/G	nd	nd	nd
OCA2	rs1800407	nd	nd	nd	G	G	G/C	G	G	G	G/C	G	G	G	nd	G	nd	nd	nd	nd
OCA2	rs1800401	C	C	nd	C	C	C	C	C	C	C/T	C	C	C	nd	C	nd	nd	C/T	nd
OCA2	rs7495174	nd	nd	nd	G	A	A	A	A	A	A	G	G	G	nd	A	G	G	nd	G/A
OCA2	rs4778241	nd	nd	nd	A	C	C	A	A	A/C	A/C	A	A	A/C	nd	C	nd	nd	C	nd
OCA2	rs4778138	nd	G	nd	nd	nd	A	nd	nd	G	G/A	nd	G	nd	A	Nd	G	G	nd	G
OCA2	rs1584407	nd	nd	nd	C	nd	A	nd	nd	C	C	C	C/A	A	C	C	C/A	C/A	nd	nd
OCA2	rs2703952	nd	A	A	A	A	A	A	A	A/G	A	C	A	A	A	A	nd	nd	nd	nd
OCA2	rs2594935	nd	nd	nd	A	G	A	G	A	A	A/G	A	A	A	nd	A	A/G	nd	A/G	nd
OCA2	rs728405	nd	nd	T	G/T	T	G/T	G	G/T	G	G/T	G	G	G	nd	G/T/A	nd	T	T	G/T
OCA2	rs1448488	A	nd	nd	A	A	A/G	G	G	A/G	A/G	nd	G	G	G	A/G	nd	A/G	nd	A/G
OCA2	rs4778220	nd	A	nd	nd	nd	A	nd	nd	A	A/G	nd	A	nd	nd	A	A	A	A	nd
OCA2	rs7170869	A	A	nd	A	A	G/A	G	G	G	G	A	G/C	G	nd	G/A	G	G	nd	G
OCA2	rs1545397	nd	A	nd	A	A	T	T	nd	A	A	nd	T	T	T	T	T	nd	nd	nd

Supporting Information Table 3. Cont

Region	SNPs ID	Subjects																		
		12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
HERC2	rs1129038	G	nd	nd	G	G/A	G/T	G	G	G/T	G	G	C/G/T	G	A	nd	nd	nd	A	nd
HERC2	rs12913832	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	nd	nd	G	nd
HERC2	rs1667394	G	nd	G	nd	nd	A	A	nd	G	A	G	G	nd	nd	nd	G/A	nd	nd	nd
HERC2	rs8039195	nd	C	nd	C	T	T	C/T	C	C	C/T	C	C	C	T	nd	C/T	nd	T	nd
HERC2	rs7183877	nd	T	C	C	C	C	C	C	C	C	C	C/A	A	C	C/A	C/A	nd	nd	nd
HERC2	rs1635168	T	nd	nd	T	G	G	T/G	T	G	T/G	T	T/G	T/G	nd	G	G	nd	G	nd
HERC2	rs8028689	C	C	nd	C	T	T	C/T	T	C/T	C/T	nd	C/T	C/T	G/T	T	nd	C	nd	nd
HERC2	rs16950987	nd	A	A	A	nd	G	nd	nd	G	A/G	nd	A/G	A	G	nd	nd	A	nd	nd
HERC2	rs916977	A	nd	nd	A	T/G	G	A/G	A	A	A/G	A	A	A	G	A	A/G	nd	G	nd
HERC2	rs7494942	nd	A	nd	A	G	G	G	A	A	A/G/C	A/C	A	A	G	A	nd	nd	G	nd
HERC2	rs3935591	nd	nd	nd	A	G	G	A/G	A	A/C	A/G	A	A	G	G	A	nd	nd	G	nd
HERC2	rs7170852	T	nd	T	T	A	A	T	T	T	T/A	T	T/G	T	A	T/A	nd	nd	A	nd
HERC2	rs2238289	nd	nd	nd	C	T	T	C	T	C/T	C/T	C	C	C	T	C	C/T	C/T	nd	C/T
HERC2	rs2240203	nd	nd	G	G	nd	A	G/A	A	G/A	G/A	G	G/A	G/A/T	nd	A	nd	G	nd	nd
HERC2	rs2240204	T	C	nd	C	C	C/A	C/T	C	C/T	C	C	C/T	C	nd	C/T	nd	T	nd	nd
HERC2	rs16950979	nd	nd	nd	A	A	A	A/G	A	A	A	A	A/G	A/G	nd	A	nd	G	nd	nd
SLC24A5 (NCKX5)	rs1426654	nd	nd	G	G	A	G	G	G	G	G	nd	G	G/A	G	G	G	G	nd	G
MC1R	rs1805007	nd	nd	nd	C	C	C	C	C	C	C	C	C	C	nd	nd	nd	nd	C/T	nd
MC1R	rs1805008	nd	nd	nd	nd	C	C/G	C	C/G	C	C	C	C	G	nd	nd	nd	nd	nd	nd
MC1R	rs3212346	A	nd	nd	A	G	G	nd	G	G	G	A	A/G	nd	G	nd	A/G	A/G	nd	nd
MC1R	rs885479	nd	nd	nd	G	G	G	G	G	G	G	G	G/A	G/A	nd	nd	G/A	nd	nd	A
DPEP1	rs164741	T	T	nd	C	C	T/G	nd	T/C	T/C	T	C	T	T/C	nd	T/C	nd	nd	nd	nd
C16orf55	rs7188458	nd	G	nd	G	G	G/A	G/A	A	G	G	A	G	A	G	G	nd	G/A	G/A	nd
C16orf55	rs459920	T	nd	nd	T	C	T	T	T	T/C	C	C	T	T	nd	T/C/G	nd	nd	nd	nd
ZNF276	rs7204478	C	nd	C	C	C	T	T	T/G	T	A/C	nd	T	T	T	T	T/C	nd	T	T
ZNF276	rs6500437	T	nd	T	T	T	C	C	C	T/C	T	nd	C	nd	C	C	T/C	nd	C	C
ZNF276	rs1800359	nd	C	C	C	C/T	C	C	C	C	C	nd	C/G	C	C	nd	C	C	C	C

Supporting Information Table 3. Cont

Region	SNPs ID	Subjects																		
		12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
ZNF778	rs9921361	nd	nd	nd	G	T	G	G	G/T	G/T	T	G/T	G/T	T	T	T	T	G/T	T	nd
PRDM7	rs2078478	T	nd	nd	T	C	C	C	C	C	C	C	T	C/T	C/T	C	T	T	nd	nd
PRDM7	rs7196459	nd	nd	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
ACSF3	rs12599126	nd	C	nd	C	T	C/T	C	C/T	C	C/T	T	C	T	nd	C/T	nd	C/T	nd	C/T
ANKRD11	rs2353033	T	nd	nd	T	C/T	C	C	C	C	C/T	C	C/T	C/T	C	T	nd	C/T	C/T	nd
ANKRD11	rs1466540	C	C	C	C	T	C	C	C	T	T	C/G	C	C	G/T	nd	C/T	nd	T	C/T
ANKRD11	rs2353028	nd	nd	nd	G	A	G/A	G	G	G	A	nd	G	nd	nd	nd	nd	nd	A	G/A
ANKRD11	rs2306633	nd	nd	A	A	G	A	A	A	A/C	G	A/G	A	A	G	A/G	A/G	nd	G	A/G
ANKRD11	rs3096304	A	nd	A	A	A	A	A	G/A	G	A	G	G/A	G	C	G/A	G/A	nd	A	A
ANKRD11	rs889574	C	nd	nd	C	C	C	C/T	C/T	C/T	C/T	nd	C	C	nd	T	nd	C/T	C	C/T
ANKRD11	rs2965946	nd	nd	nd	T	C	C	C/T/A	C	C	C/T	C	C	C	T	C	C	C	C	nd
SPG7	rs382745	nd	nd	nd	C/T	C/T	C	C/T	T	C	C	C	C	T	nd	C/T	C	C/T	C/T	nd
CPNE7	rs455527	nd	nd	T	T	T	T	C	T	T	T	T	T/C	T/C	C	A/C	nd	nd	nd	T/C
CPNE7	rs352935	G	G	nd	G	G/A	nd	G	G	G	G	G	G	G	G	G	nd	G/A	G/A	nd
CPNE7	rs464349	nd	nd	nd	T	T/C	A	nd	C	C	C	C	T/C	T/C	nd	T/C	C	T/C	T/C	C
CHMP1A	rs460879	T	T	nd	T	T	C	nd	C	C/T	C	C	C/T	C	C	C/T	C/T	nd	nd	nd
CDK10	rs258322	nd	T	T	T	C	C/T	C	C	C	T	nd	C/T	T	C	C	C/T	C/T	C/T	C/T
CDK10	rs258324	nd	C	C	C	C	C/A	C/A	C	C	C/A	nd	C	C	A	nd	C/A	C	C	C/A
CDK10	rs3751700	G	nd	nd	G	G	G/A	G/A	G	G	G/C	G	G/A	G	A	G/A	G/A	nd	nd	G/A
CDK10	rs1946482	T	T	nd	T	T	T/C	T/C	T	T	T	T	T/C	T	nd	C	T/C	nd	nd	T/C
SPATA2L	rs3751695	nd	nd	C	C	C	T	C	T	C/T	C	nd	C/T	C	nd	nd	C/T	nd	nd	C/T
FANCA	rs7195066	G	A	nd	A	A/G	A	A	A	A/G	A/G	G	A	A	A	G	A	A	A/G	A
FANCA	rs8058895	T	nd	nd	T	T/A	T	T	T	T	T	T	T	T	nd	T	nd	nd	nd	nd
FANCA	rs2011877	nd	nd	nd	A	A/C	C	C	C	C	A/C	A/C	C	C	nd	A/C	A/C	nd	C	C
FANCA	rs2239359	A	A	A	A	nd	A/G	nd	A	G	A	A	A/G	A/G	nd	A/G	A	A	nd	A/G
FANCA	rs16966142	C	nd	nd	C	C	C/T	nd	T	C/G	C/T	C	C/T	C	T	C	C/T	nd	nd	C/T
FANCA	rs1800286	nd	nd	nd	G	A	G	G	G	G	G	nd	G	G	nd	G/A	G	G	G	G

Supporting Information Table 3. Cont

Region	SNPs ID	Subjects																		
		12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
FANCA	rs11861084	nd	nd	nd	C	A	nd	C	C	C	C	A	C	C	C	C/A	C	C	C	C
SPIRE2	rs8060934	nd	nd	nd	C	C/T	C/T	T	T	T	C/T	C	C/T	nd	nd	T	C/T	C/T	T	T
SPIRE2	rs3803688	T	T	T	nd	T	T/C	T/C	nd	C	T	C	T/C	C	C	T/C	T/C	T/C	nd	nd
TCF25	rs2270460	nd	T	nd	T	G	G	G	nd	G	G	nd	G	G	nd	G	nd	G/T	nd	nd
CENPBD1	rs4785755	nd	C	nd	C	T	T	T	T	C/T	C/T	C	C/T	C/A	T	A/T	C/T	nd	T	T
DBNDD1	rs8059973	nd	G	nd	G	G	A	G	A	G	G/A	A	G	G	A	A	G/A	nd	G	G/A
DBNDD1	rs11648785	C	T	C	C	C/T	C	C	C	C	C/T	C/T	C/T	C/T	C	C/T	nd	nd	C/T	nd
GAS8	rs2241039	nd	C	C	C	nd	T	T	T	T/C	T	T	T/C	T/C	T	T/A	T/C	T	T/C	nd
GAS8	rs3785181	G	nd	G	G	G	G/A	nd	A	G	G	G	G	G	G/A	G/A	G/A	A	nd	nd
GAS8	rs1048149	nd	nd	nd	C	C	C	C	T	C/T	C/T	nd	C	nd	C	C	nd	nd	nd	nd
AFG3L1	rs4785763	nd	nd	nd	A	C	nd	nd	C	A/C	C	G	A/C	A	C	nd	A/C	A/C	A/C	A/C
AFG3L1	rs4408545	nd	C	nd	nd	T	C/T	C	C	C	C/T	nd	C/T	C	C	C/T	nd	nd	nd	nd
DYNLRB1	rs2281695	C	C	nd	nd	C	A/C/G	C	C	C	A/C	C	C	C	nd	nd	C	C	C	T/C
PIGU	rs2378199	C	nd	C	C	T	C	C	C	C	C	C	C	C	C	T/C	C	C	C	T/C
PIGU	rs2378249	nd	A	nd	A	A/G	A	A	A	A	A/C	A	A	A	A	A	A	A	A	A/G
NCOA6	rs6060034	nd	nd	nd	nd	C	C	C	C	C	C	nd	C	C	nd	C/T	C	C/T	C	C/T
NCOA6	rs6060043	nd	nd	T	T	T/C	T/G	T	T	T/G	T/G	T	T	T	nd	T/C	T	T/C	T	T/C
EIF6	rs619865	G	G	nd	G/A	G	G	nd	G	G	G	G	G	G	G	G	G	G	G	G
ASIP	rs6058017	nd	nd	nd	G	G/A	A	A	A	A	A	nd	A	G	nd	nd	nd	G/A	nd	nd
Intergenic regions	rs9328192	nd	nd	A	A	A/G	A/G	G	A/G	A/G	A	G	A	A	nd	A	nd	nd	G	nd
	rs9405681	T	T	T	T	C/G	C/T	C	C/T	C	C/T/G	C	C/T	C/G	nd	C	C	C	C	C
	rs4959270	C	C	nd	A	C/A	G/A	A	C/A	C	A	C	C	C	A	C	C/A	nd	A	nd
	rs9378805	nd	nd	A	A	A/C	C	C	nd	A	A	A	A	A/C	nd	A	nd	nd	C	A/C
	rs1540771	A	A	A	A	G	G/A	A	G/A/C	G/C	G	G	G	nd	A	G	G/A	nd	A	G/A
	rs1011176	G	nd	nd	G	G/A	A	A	nd	G/A	A	G	G/A	G/A	A	A	nd	nd	G	G/A
	rs2305498	nd	nd	nd	C	C	G/T	C	C/T	C/T	C/T	C	C	G/T	C/T/G	C	C/T	nd	nd	nd
	rs12821256	nd	nd	nd	T	T	T	T	nd	T	T	T	T	T	T	T	nd	nd	C	nd

Supporting Information Table 3. Cont

Region	SNPs ID	Subjects																		
		12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Intergenic regions	rs8016079	nd	G	nd	G	G	G	G	G/A	G	G	G	G/A	G	G	G	nd	nd	nd	nd
	rs4904864	G	nd	G	G	G	G/A	nd	A	G/A	G	nd	G/A	G	A	nd	G	nd	nd	nd
	rs4904868	nd	nd	nd	T	T/C	T/C	C	T	C	T/C	T	T/C	C	nd	T	C	nd	T/C	nd
	rs12896399	nd	nd	G	G	G	G/T	T	T	G	G	G	G	T	nd	G	T	nd	G/T	nd
	rs438702	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
	rs9932354	T	nd	nd	nd	A	C	C/A	C/A/G	C/G	C	nd	C	C	C	C/A	nd	nd	C/A	nd
	rs11076747	G	G	G	G	nd	G	G	G	G	G	G	G	G	G	G/A	nd	nd	G/A	nd
	rs4785648	nd	nd	A	A	G	A/G	G	G	A	G	A/G	A/G	A	nd	A/G	G	nd	G	G
	rs4347628	nd	nd	C	C	C/T	C	T	C/T	C/T	C/T	T	C	C/T	T	C/T	C	C/T	C	nd
	rs12443954	G	nd	G	G	A	A	nd	A	G/A/C	A	A	G	A	G	G/A	G	G/A	nd	G/A
	rs9936896	nd	C	C	C	T	T	T	T	C	T	C/T	C/T	C	C	T	C/T	nd	nd	C/T
	rs4785612	nd	nd	nd	C	A	A	A	nd	C/A	C	A	C/A	nd	nd	A	C/A	A	A	nd
	rs7201721	nd	A	nd	A	A	A	A	A	A	A/G	A	A	A	A	A	A	A	A	A
	rs4238833	nd	G	nd	G	T	T	T	T	G/T	G/T	G	G/T	G	T	T	G/T	G/T	G/T	T
	rs6119471	nd	nd	nd	G	C	C	C	C	G/C	C	A/C	C	nd	C	C	nd	nd	nd	nd
	rs1015362	nd	nd	nd	A	G	G	nd	G	A	A/G	A	A/G	G	G	A	nd	A/G	nd	A/G
	rs4911414	nd	nd	nd	G	G	G	nd	G	G	G/T	G	G/T/A	G	G	G/T	G	G/T	G	nd

12= Neanderthal (Vi33.16); 13= Neanderthal (Vi33.25); 14= Neanderthal (Vi33.26); 15= Denisova; 16= Sardinian (HGDP00665); 17= Papuan (HGDP00542); 18= Papuan (HGDP00551); 19= Melanesian (HGDP00491); 20= San (HGDP01029); 21= Yoruba (HGDP00927); 22= Mbuti Pygmy (HGDP00456); 23= Han (HGDP00778); 24= Cambodian (HGDP00711); 25= Native American (HGDP00998); 26= Mongolian (HGDP01224); 27= Han Chinese Individual (YanHuang Project); 28= Anonymous Korean Individual, AK1 (Genomic Medicine Institute); 29= Anonymous Irish Male; 30= Extinct Paleo-Eskimo (Saqqaq Genome Project); A= Adenine; *ACSF3* = acyl-CoA synthetase family member 3; *AFG3L1* = ATPase family gene 3 like 1 (Yeast); *ANKRD11* = ankyrin repeat domain 11; *ASIP* = agouti signaling protein; C= Cytosine; *C16orf55* = chromosome 16 open reading frame 55; *CDK10* = cyclin-dependent kinase 10; *CENPBD1* = CENPB DNA-binding domains containing 1; *CHMP1A* = chromatin modifying protein 1A; *CPNE7* = copine VII; *DBNDD1* = dysbindin (dystrobrevin binding protein 1) domain containing 1; *DPEP1* = dipeptidase 1; *DYNLRB1* = dynein, light chain, roadblock-type 1; *EIF6* = eukaryotic translation initiation factor 6; *FANCA* = Fanconi anemia, complementation group A; G= Guanine; *GAS8* = growth arrest-specific 8; *HERC2* = hect domain and RLD 2; *IRF4* = interferon regulatory factor 4; *MATP* = membrane associated transporter protein; *MC1R* = melanocortin 1 receptor; *NCKX5*= Na/Ca/K exchanger 5; *NCOA6* = nuclear receptor coactivator 6; nd = non-determined; *OCA2* = oculocutaneous albinism II; *PIGU* = phosphatidylinositol glycan anchor biosynthesis, class U; *PRDM7* = PR domain containing 7; *SLC24A4* = solute carrier family 24, member 4; *SLC24A5* = solute carrier family 24, member 5; *SLC45A2* = solute carrier family 45 member 2; *SPATA2L* = spermatogenesis associated 2-like; *SPG7* = spastic paraplegia 7; *SPIRE2* = spire homolog 2 (Drosophila); T= Timine; *TCF25* = transcription factor 25; *TPCN2* = two pore segment channel 2; *TYR* = tyrosinase; *ZNF276* = zinc finger protein 276; *ZNF778* = zinc finger protein 778.

Supporting Information Table 4. Possible phenotypes of carriers of the indicated SNPs

Region	SNPs ID	Mutation	Possible Phenotypes	References
SLC45A2 (MATP)	rs26722	Glu272Lys - G>A	G= fairer skin; lighter brown hair; not-brown eyes/ A= darker skin; darker brown hair; brown eyes	[1]
SLC45A2 (MATP)	rs6867641	C>T	C= fairer skin/ T=darker skin	[2]
IRF4	rs12203592	C>T	T=green or blue eyes; fairer skin/ C=brown or black eyes; darker skin	[3, 4, 5, 6]
TPCN2	rs3829241	Gly734Glu - G>A	A= blond or red hair; fairer skin/ G= not-blond or not-red hair; darker skin	[7]
TPCN2	rs3750965	Lys376Arg - A>G	A = blond or red hair/ G= brown hair	[7]
TPCN2	rs35264875	Met484Leu - A>T	A= brown hair; not-blond hair; darker skin/ T= not-brown hair; blond hair; fairer skin	[7]
TPCN2	rs896978	C>T	T=brown hair/ C= not-brown hair	[7]
TYR	rs1042602	Ser192Tyr - C>A	C= darker skin; freckles; brown hair/ A= fairer skin; not-freckles; not-brown hair	[7, 8]
TYR	rs1126809	Arg402Gln - G>A	G= not-blue eyes; brown hair/ A=blue eyes; not-brown hair	[7]
TYR	rs1393350	G>A	G= green or brown or black eyes/ A= blue eyes	[4, 6, 7, 9]
SLC24A4	rs2402130	G>A	G=green eyes; brown hair/ A= not-green eyes; not-brown hair	[7]
OCA2	rs1498519	C>A	C= brown eyes/ A=not-brown eyes	[7]
OCA2	rs1800414	His615Arg - A>G	A= darker skin/ G= fairer skin	[10]
OCA2	rs1800407	Arg419Gln - G>A	G= not-green or brown or black eyes/ A= blue eyes	[6, 11]
OCA2	rs1800401	Arg305Trp - C>T	C= blue eyes/ T= not-blue eyes	[11]
OCA2	rs7495174	G>A	A = blue eyes; lighter brown or blond hair; fairer skin/ G= green or brown eyes; darker brown or black hair; darker skin	[3, 4, 7, 12, 13, 14]
OCA2	rs4778241	A>C	C= blue eyes; lighter brown or blond hair; fairer skin/ A= not-blue eyes; darker brown or black hair; darker skin	[4, 7, 12, 13, 14, 15]
OCA2	rs4778138	G>A	G= green or brown eyes; darker brown hair/ A= blue eyes; fairer skin; lighter brown hair	[4, 7, 12, 13, 14]
OCA2	rs1584407	C>A	C= blue eyes/ A= not-blue eyes	[7]
OCA2	rs2703952	A>C	C= green or brown eyes/ A= not-green or not-brown eyes	[7]
OCA2	rs2594935	A>G	G= blue eyes/ A= not-blue eyes	[7]
OCA2	rs728405	G>T	T= blue eyes/ G= not-blue eyes	[7]
OCA2	rs1448488	A>G	G= green or brown eyes/ A= not-green or not-brown eyes	[7]
OCA2	rs4778220	A>G	G= green or brown eyes/ A= not-green or not-brown eyes	[7]
OCA2	rs7170869	G>A	G= green or brown eyes/ A= not-green or not brown eyes	[7]

Supporting Information Table 4. Cont

Region	SNPs ID	Mutation	Possible Phenotypes	References
OCA2	rs1545397	A>T	T=fairer skin; not-brown eyes/ A= darker skin; brown eyes	[5]
HERC2	rs1129038	G>A	A= blue eyes/ G=brown eyes	[12, 15]
HERC2	rs12913832	A>G	G= fairer skin; blue or not-brown eyes; lighter brown hair/ A= not-blue or brown eyes, darker skin; darker brown hair	[4, 5, 6, 12, 15, 16, 17]
HERC2	rs1667394	G>A	A= blue eyes; blond hair/ G= green or brown eyes; brown hair	[4, 9, 12]
HERC2	rs8039195	C>T	T= blue eyes; blond hair/ C= not-blue eyes; not-blond hair	[4, 7]
HERC2	rs7183877	C>A	C= blue eyes; blond hair/ A= not-blue eyes; not-blond hair	[4, 7]
HERC2	rs1635168	T>G	T= green or brown eyes/ G= not-green or not-brown eyes	[4, 7]
HERC2	rs8028689	C>T	T= blue eyes/ C= not-blue eyes	[4, 7]
HERC2	rs16950987	A>G	G= blue eyes/ A= not-blue eyes	[4, 7]
HERC2	rs916977	A>G	G= blue eyes/ A= brown eyes	[12, 13, 15]
HERC2	rs7494942	A>G	G= blue eyes/ A= brown eyes	[12]
HERC2	rs3935591	A>G	G= blue eyes/ A= brown eyes	[15]
HERC2	rs7170852	T>A	A= blue eyes/ T=brown eyes	[15]
HERC2	rs2238289	C>T	T=blue eyes/ C=brown eyes	[15]
HERC2	rs2240203	G>A	A= blue eyes/ G= brown eyes	[15]
HERC2	rs2240204	C>T	T= green or brown eyes/ C= not-green or not-brown eyes	[7]
HERC2	rs16950979	A>G	G= green or brown eyes/ A= not-green or not-brown eyes	[7]
SLC24A5 (NCKX5)	rs1426654	Ala111Thr - G>A	G= more melanin/ A=lesser melanin	[5, 18]
MC1R	rs1805007	Arg151Cys - C>T	T= blond or red hair; freckles/ C= not-blond or not-red hair; not-freckles	[7, 9]
MC1R	rs1805008	Arg160Trp - C>T	T= blond or red hair; freckles/ C= not-blond or not-red hair; not-freckles	[7, 9]
MC1R	rs3212346	A>G	G= red hair/ A= not-red hair	[7]
MC1R	rs885479	Arg163Gln - G>A	G= red hair; darker skin; not-brown eyes/ A= fairer skin; brown eyes; not-red hair	[5, 7, 19]
DPEP1	rs164741	T>C	T= red hair; freckles/ C= not-red hair; not-freckles	[3, 4, 7]
C16orf55	rs7188458	G>A	G= not-red hair; not-freckles/ A= red hair; freckles	[3, 4, 7]
C16orf55	rs459920	T>C	T= red hair; freckles/ C= not-red hair; not-freckles	[7]
ZNF276	rs7204478	T>C	T= red hair; freckles/ C= not-red hair; not-freckles	[3, 4, 7]
ZNF276	rs6500437	Trp188Arg - T>C	T= not-red hair/ C= red hair	[7]
ZNF276	rs1800359	C>T	T= not-red hair; not-freckles/ C= red hair; freckles	[7]

Supporting Information Table 4. Cont

Region	SNPs ID	Mutation	Possible Phenotypes	References
ZNF778	rs9921361	Gln553His - G>T	T= red hair/ G= not-red hair	[7]
PRDM7	rs2078478	C>T	T= not-red hair/ C= red hair	[7]
PRDM7	rs7196459	G>T	T= red hair; freckles/ G= not-red hair; not-freckles	[4, 7]
ACSF3	rs12599126	C>T	T= red hair/ C= not-red hair	[7]
ANKRD11	rs2353033	C>T	T= not-freckles; not-red hair/ C= freckles; red hair	[4, 7]
ANKRD11	rs1466540	C>T	T= red hair/ C= not-red hair	[7]
ANKRD11	rs2353028	G>A	G= not-red hair/ A= red hair	[7]
ANKRD11	rs2306633	A>G	G= red hair/ A= not-red hair	[7]
ANKRD11	rs3096304	G>A	G= not-red hair/ A= red hair	[7]
ANKRD11	rs889574	C>T	T= freckles/ C= not-freckles	[7]
ANKRD11	rs2965946	C>T	T= freckles/ C= not-freckles	[7]
SPG7	rs382745	C>T	T= red hair; freckles/ C= not-red hair; not-freckles	[7]
CPNE7	rs455527	Phe77Leu - T>C	C= not-red hair/ T= red hair	[7]
CPNE7	rs352935	G>A	G= freckles; red hair/ A= not-freckles; not-red hair	[7]
CPNE7	rs464349	T>C	T= not-freckles; not-red hair/ C= freckles; red hair	[7]
CHMP1A	rs460879	C>T	T= not-freckles; not-red hair/ C= freckles; red hair	[7]
CDK10	rs258322	C>T	T= red hair; freckles/ C= not-red hair; not-freckles	[3, 4, 7]
CDK10	rs258324	C>A	C= red hair/ A= not-red hair	[7]
CDK10	rs3751700	G>A	G= red hair/ A= not-red hair	[7]
CDK10	rs1946482	T>C	T= red hair/ C= not-red hair	[7]
SPATA2L	rs3751695	Phe156Phe - C>T	T= red hair; freckles/ C= not-red hair; not-freckles	[7]
FANCA	rs7195066	Asp809Gly - A>G	A= not-red hair; not-freckles/ G= red hair; freckles	[4, 7]
FANCA	rs8058895	T>C	T= not-red hair; not-freckles/ C= red hair; freckles	[7]
FANCA	rs2011877	A>C>T	C= red hair	[7]
FANCA	rs2239359	Ser501Gly - A>G	A= freckles/ G= not-freckles	[7]
FANCA	rs16966142	C>T	T= not-red hair/ C= red hair	[7]
FANCA	rs1800286	G>A	G= red hair; freckles/ A= not-red hair; not-freckles	[7]
FANCA	rs11861084	C>A	C= red hair; freckles/ A= not-red hair; not-freckles	[7]

Supporting Information Table 4. Cont

Region	SNPs ID	Mutation	Possible Phenotypes	References
SPIRE2	rs8060934	C>T	T= red hair; freckles/ C= not-red hair; not-freckles	[7]
SPIRE2	rs3803688	T>C	T= red hair/ C= not-red hair	[7]
TCF25	rs2270460	G>T	T= not-red hair/ G= red hair	[7]
CENPBD1	rs4785755	Thr168Ile - C>T	C= red hair; freckles/ T= not-red hair; not-freckles	[7]
DBNDD1	rs8059973	G>A	G= red hair/ A= not-red hair	[7]
DBNDD1	rs11648785	C>T	T= not-red hair; not-freckles/ C= red hair; freckles	[7]
GAS8	rs2241039	T>C	T= not-red hair; not-freckles/ C= red hair; freckles	[3, 4, 7]
GAS8	rs3785181	G>A	G= red hair/ A= not-red hair	[7]
GAS8	rs1048149	C>T	T= red hair; freckles/ C= not-red hair; not-freckles	[7]
AFG3L1	rs4785763	A>C	A= blond or red hair; freckles/ C= not-blond or not-red hair; not-freckles	[3, 7]
AFG3L1	rs4408545	C>T	T= not-red hair; not-freckles/ C= red hair; freckles	[3, 4, 7]
DYNLRB1	rs2281695	T>C	T= freckles/ C= not-freckles	[7]
PIGU	rs2378199	T>C	T= freckles/ C= not-freckles	[7]
PIGU	rs2378249	A>G	G= freckles/ A= not-freckles	[7]
NCOA6	rs6060034	C>T	T= freckles/ C= not-freckles	[7]
NCOA6	rs6060043	T>C	T= not-freckles/ C= freckles	[7]
EIF6	rs619865	G>A	G= not-freckles/ A= freckles	[7]
ASIP	rs6058017	G>A	G= more melanin; not-freckles/ A= lesser melanin; freckles	[7, 20, 21]
Intergenic regions	rs9328192	A>G	G= freckles/ A= not-freckles	[7]
	rs9405681	C>T	T= not-freckles/ C= freckles	[7]
	rs4959270	C>A	C= not-freckles/ A= freckles	[7]
	rs9378805	A>C	C= freckles/ A= not-freckles	[7]
	rs1540771	G>A	A= brown hair; freckles/ G= not-brown hair; not-freckles	[4, 7, 9]
	rs1011176	G>A	A= blond hair/ G= brown hair	[7]
	rs2305498	C>T	T= blond hair/ C= not-blond hair	[7]
	rs12821256	T>C	C= blond hair/ T= brown hair	[4, 7]
	rs8016079	G>A	G= blue eyes/ A= not-blue eyes	[7]
	rs4904864	G>A	G= blond hair; blue eyes/ A= not-blond hair; not-blue eyes	[7]

Supporting Information Table 4. Cont

Region	SNPs ID	Mutation	Possible Phenotypes	References
Intergenic regions	rs4904868	T>C	T= brown hair; not-blue eyes/ C= not-brown hair; blue eyes	[7]
	rs12896399	G>T	T= blue eyes; blond hair/ G= brown or black eyes; not-blond hair	[4, 6, 7, 19]
	rs438702	G>A	A= blue eyes/ G= not-blue eyes	[7]
	rs9932354	C>A	C= not-red hair/ A= red hair	[7]
	rs11076747	G>A	G= not-red hair/ A= red hair	[7]
	rs4785648	A>G	G= red hair/ A= not-red hair	[7]
	rs4347628	C>T	T= not-red hair/ C= red hair	[7]
	rs12443954	G>A	G= not-red hair/ A= red hair	[7]
	rs9936896	C>T	T= not-red hair; not-freckles/ C= red hair; freckles	[7]
	rs4785612	C>A	C= red hair; freckles/ A= not-red hair; not-freckles	[7]
	rs7201721	A>G	G= red hair/ A= not-red hair	[7]
	rs4238833	G>T	T= not-red hair; not-freckles/ G= red hair; freckles	[3, 4, 7]
	rs6119471	G>C	G=brown eyes; darker skin/ C= not-brown eyes; fairer skin	[5]
	rs1015362	A>G	G= freckles; red or blond hair/ A= not-freckles; not-red or not-blond hair	[7]
	rs4911414	G>T	T= freckles; red or blond hair/ G= not-freckles; not-red or not-blond hair	[7]

A= Adenine; *ACSF3* = acyl-CoA synthetase family member 3; *AFG3L1* = ATPase family gene 3 like 1 (Yeast); Ala = alanine; *ANKRD11* = ankyrin repeat domain 11; Arg = arginine; *ASIP* = agouti signaling protein; Asp = aspartic acid; C= Cytosine; *C16orf55* = chromosome 16 open reading frame 55; *CDK10* = cyclin-dependent kinase 10; *CENPBD1* = CENPB DNA-binding domains containing 1; *CHMP1A* = chromatin modifying protein 1A; *CPNE7* = copine VII; Cys = cysteine; *DBNDD1* = dysbindin (dystrobrevin binding protein 1) domain containing 1; *DPEP1* = dipeptidase 1; *DYNLRB1* = dynein, light chain, roadblock-type 1; *EIF6* = eukaryotic translation initiation factor 6; *FANCA* = Fanconi anemia, complementation group A; G= Guanine; *GAS8* = growth arrest-specific 8; Gln = glutamine; Glu = glutamine acid; Gly = glycine; *HERC2* = hect domain and RLD 2; His = histidine; Ile = isoleucine; *IRF4* = interferon regulatory factor 4; Leu = leucine; Lys = lysine; *MATP* = membrane associated transporter protein; *MC1R* = melanocortin 1 receptor; Met = methionine; *NCKX5*= Na/Ca/K exchanger 5; *NCOA6* = nuclear receptor coactivator 6; nd = non-determined; NS = non-synonymous; *OCA2* = oculocutaneous albinism II; Phe = phenylalanine; *PIGU* = phosphatidylinositol glycan anchor biosynthesis, class U; *PRDM7* = PR domain containing 7; S = synonymous; Ser = serine; *SLC24A4* = solute carrier family 24, member 4; *SLC24A5* = solute carrier family 24, member 5; *SLC45A2* = solute carrier family 45 member 2; *SPATA2L* = spermatogenesis associated 2-like; *SPG7* = spastic paraplegia 7; *SPIRE2* = spire homolog 2 (Drosophila); T= Timine; *TCF25* = transcription factor 25; Thr = threonine; *TPCN2* = two pore segment channel 2; Trp = tryptophan; Tyr = tyrosine (amino acid); *TYR* = tyrosinase; UTR = untranslated region; *ZNF276* = zinc finger protein 276; *ZNF778* = zinc finger protein 778.

SUPPORTING INFORMATION REFERENCES

[1] Graf J, Hodgson R, van Daal A. 2005. Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mut 25:278-284.

[2] Graf J, Voisey J, Hughes I, van Daal, A. 2007. Promoter polymorphisms in the MATP (SLC45A2) gene are associated with normal human skin color variation. Hum Mut 28:710-717.

[3] Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, Hu FB, Thomas G, Hoover RN, Chanock S, Hunter DJ, Han J. 2009. Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol 129: 2250–2257.

[4] Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ. 2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4:e1000074.

[5] Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, Caragine T, Prinz M, Wurmbach E. 2011. Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Sci Int Genet 5:472-478.

[6] Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, Kayser M, Ballantyne KN. 2011. Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5:464-471.

[7] Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G, Pálsson S, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Vermeulen SH, Goldstein AM, Tucker MA, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K. 2008. Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40:835-837.

[8] Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA. 2003. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112:387–399.

[9] Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, Jakobsdottir M, Steinberg S, Pálsson S, Jonasson F, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K. 2007. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39: 1443-1452.

[10] Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, Jin L, Parra EJ. 2010. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: Further evidence of convergent evolution of skin pigmentation. PLoS Genet 6:e1000867.

[11] Rebbeck TR, Kanetsky PA, Walker AH, Holmes R, Halpern AC, Schuchter LM, Elder DE, Guerry D. 2002. P gene as an inherited biomarker of human eye color. Cancer Epid Biomarkers Prev 11:782–784.

[12] Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, Martin NG, Montgomery GW. 2008. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet 82:424-31.

[13] Kayser M, Liu, F, Janssens AC, Rivadeneira F, Lao O, Van Duijn K, Vermeulen M, Arp P, Jhamai MM, Van Ijcken WF, Den Dunnen JT, Heath S, Zelenika D, Despriet DD, Klaver CC, Vingerling JR, De Jong PT, Hofman A, Aulchenko YS, Uitterlinden AG, Oostra BA, Van Duijn CM. 2008. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet 82:411-423.

[14] Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, Hayward NK, Martin NG, Sturm RA. 2007. A three single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am J Hum Genet 80:241–252.

[15] Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L. 2008. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet 123:177-187.

[16] Bouakaze C, Keyser C, Crubézy E, Montagnon D, Ludes B. 2009. Pigment phenotype and biogeographical ancestry from ancient skeletal remains: Inferences from multiplexed autosomal SNP analysis. Int J Legal Med 123:315-25.

[17] Branicki W, Brudnik U, Wojas-Pelc A. 2009. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. Ann Hum Genet 73:160-170.

[18] Sturm RA. 2009. Molecular genetics of human pigmentation diversity. Hum Mol Genet 18:R9-R17.

[19] Póspiech E, Draus-Barini J, Kupiec T, Wojas-Pelc A, Branicki W. 2011. Gene-gene interactions contribute to eye colour variation in humans. J Hum Genet 56:447-55.

[20] Bonilla C, Boxill LA, Donald SA, Williams T, Sylvester N, Parra EJ, Dios S, Norton HL, Shriver MD, Kittles RA. 2005. The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. Hum Genet 116:402–406.

[21] Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR. 2002. A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet 70:770–775.

3.2. CAPÍTULO II

“Implications of the miscegenation process in skin color molecular assessment**”**

(Artigo submetido na revista *PLoS One*, 2013)

Implications of the miscegenation process in skin color molecular assessment

C.C.S. Cerqueira[a], T. Hünemeier[a], J.A. Gomez-Valdes[b], V. Ramallo[a], C.D. Volasko-Krause[a], A.A.L. Barbosa[c], P. Vargas-Pinilla[a], R.C. Dornelles[a], D. Longo[d], F. Rothhammer[e], G. Bedoya[f], S. Canizales-Quinteros[g,h], V. Acuña-Alonzo[i], C. Gallo[j], G. Poletti[j], R. González-José[k], F.M. Salzano[a], L. Schuler-Faccini[a,l], A. Ruiz-Linares[m], M.C. Bortolini[a,*] for CANDELA (*Consortium for the Analysis of the Diversity and Evolution of Latin America)*.

[a]Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[b]Laboratorio de Antropología Física, Departamento de Anatomía, Facultad de Medicina, Universidad Nacional Autónoma de Mexico,Mexico City, Mexico

[c]Departamento de Ciências Biológicas, Universidade Estadual do Sudoeste da Bahia, Jequié, Brasil

[d]Instituto Federal de Educação, Ciência e Tecnologia Farroupilha, Alegrete, Brazil

[e]Instituto de Alta Investigación, Universidad de Tarapacá, Facultad de Medicina, Universidad de Chile and Centro de Investigaciones del Hombre en el Desierto, Arica, Chile

[f]Universidad de Antioquia, Medellín, Colombia

[g]Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, Mexico

[h] Instituto Nacional de Medicina Genómica, Mexico City, Mexico

[i]Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia, Mexico City, Mexico

[j]Laboratorio de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú

[k]Centro Nacional Patagónico, CONICET, Puerto Madryn, Argentina

[l]INAGEMP – Instituto Nacional de Genética Médica Populacional, Porto Alegre, Brazil

[m]Department of Genetics, Evolution and Environment and UCL Genetics Institute, University College London, London, UK

*Corresponding author: Tel.: +55 51 3308 9844. *E-mail address*: maria.bortolini@ufrgs.br

ABSTRACT

The understanding of the complex genotype-phenotype architecture of human pigmentation has clear implications for the evolutionary history of humans, as well as for medical and forensic practices. Although dozens of genes have previously been associated with human skin color, knowledge about this trait remains incomplete. In particular, studies focusing on populations outside the European-North American axis are rare, and, until now, admixed populations are seldom considered. The present study was designed to help fill this gap. Our objective was to verify possible associations of 18 single nucleotide polymorphisms (SNPs), located within 9 genes, and one pseudogene with the Melanin Index (MI) in two admixed Brazilian populations (*Gaucho*, N = 354; *Baiano*, N = 149) with different histories of geographic and ethnic colonization. Some of these polymorphisms have never been studied in the context of normal variation in human skin color. Of the total sample, 8 markers were found to be associated with skin color (4 in *Gaucho*, 6 in *Baiano*), but only 3 (*HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982) showed a statistical significance in both samples. Therefore, only these 3 SNPs should be preliminarily considered to have forensic significance because they consistently showed the association independently of the population considered. These results also demonstrate the importance of considering possible population differences in selecting an appropriate set of SNPs as phenotype predictors in forensic practice.

Key words: Admixed populations, Forensic DNA phenotyping, Human skin color

1. Introduction

In humans and other animals, pigmentation results from melanin synthesis, which occurs in melanocytes stored in vesicles known as melanosomes. Melanocytes are found in the epidermis, hair bulb, and iris, and are directly responsible for the pigmentation of these organs and structures [1, 2]. The color spectrum observed in the human eyes, hair, and skin is related to the type and quantity of melanin produced, the pattern of packaging, size, and number of melanosomes, as well as to the pH of the melanosomal environment [3, 4]. There are 2 main melanin types: eumelanin and pheomelanin, which are responsible for the black/brown color and the yellow/red spectrum, respectively.

The human pigmentation pathway is regulated by a network controlled by many genes, as well as by environmental, mechanical, and epigenetic factors. For instance, Quillen et al. [5] identified 76 skin pigmentation candidate genes from the Online Mendelian Inheritance in Man (OMIM) database using the following key-words: ‘‘color’’ and ‘‘pigmentation’’. The gene set identified included both classical and non-classical genes. The pigmentation candidate genes identified to date act over different stages of the above-mentioned processes [6]. Additional information about the function, localization, number of exons, and other relevant data related to these and others genes can be found in Cerqueira et al. [7, 8].

A considerable number of genetic studies involving normal and pathological variation in human pigmentation have emerged recently [5, 8-19]. Some of these investigations showed that it is possible to predict human color phenotypes from genotype data with different levels of accuracy, thus highlighting the importance of these types of studies for forensic practices [8, 16]. Furthermore, it was suggested that some variants associated to pigmentation also confer susceptibility or resistance to skin cancer [12, 20, 21], as the presence of several skin tones are attributed to adaptation to diverse environments after dispersal of the anatomically modern *Homo sapiens* from Africa to other continents [9, 13-15, 22-25].

An additional aspect that emerged from these studies is the diversity of the pigmentation genotype-phenotype architecture across human populations. For example, Quillen et al. [5] showed that in addition to the classical genes *SLC24A5* and *SLC45A2*, others, such as *OPRM1* and *EGFR*, have also contributed to differences in pigmentation between Native Americans and Europeans. Moreover,

Norton et al. [6] suggested that polymorphisms in *SLC24A5*, *SLC45A2*, and *TYR* had a predominant role in the evolution of lighter skin color in Europeans, but not in East Asians, indicating recent convergent evolution of lighter pigmentation phenotypes, and the possible role of natural selection in this process. Quillen and Shriver [15] evaluated skin color architecture at the population level and summarised this situation as follows: (a) despite gene flow among human populations, differences in allele frequencies at pigmentation genes are nonetheless observed, and (b) signatures of natural selection in some of these genes are observed among human populations, whereas in other genes, the signal is population-exclusive.

As noted above, several studies have indicated different genetic backgrounds of pigmentation genes in different populations. Therefore, caution should be exercised when extrapolating results from one population to another. The gene pool of contemporary admixed Brazilian populations is characterised by different levels of contributions from their continental ancestral groups, Native American, European, and African, and striking differences have been reported among them [26, 27]. Nonetheless, few studies have examined polymorphisms in pigmentation candidate genes in these Brazilian populations to date [28]. Thus, the main objective of the present work was to evaluate the possible association of 18 single nucleotide polymorphisms (SNPs), located in 9 pigmentation candidate genes and one pseudogene, with the Melanin Index (MI) in 2 Brazilian populations with different admixture levels.

2. Materials and methods

2.1. Recruitment of study samples

This study recruited 503 Brazilian volunteers of the Consortium for the Analysis of the Diversity and Evolution of Latin America (CANDELA, http://www.ucl.ac.uk/silva/candela). They were recruited in Porto Alegre, the capital of the southernmost state of Brazil, Rio Grande do Sul (RS), and in Jequié, Bahia state (BA), whose effective Portuguese colonization date is much earlier than that of RS.

The states indicated above have distinct colonial histories; RS is characterised by more substantial European colonization (a significant number of non-Portuguese

European immigrants arrived in RS in the 19th century), whereas African heritage is more visible in BA with respect to both phenotypic and cultural aspects. Apart from these widely recognised characteristics, other historical, demographic, and genetic particularities can distinguish these populations [26, 27, 29-33, and references therein].

The traditional *Gaucho* culture has roots in the South American Pampa region, whose geographical areas cover parts of RS, Argentina, and Uruguay. However, in modern Brazil, the word *Gaucho* (*Gaúcho* in Portuguese) is used to refer to anyone born in RS, and not only to those born in the Pampean region of this state. In the present study, we employ this term to refer to any volunteer from RS, whereas the term *Baiano* is used to identify individuals from BA.

This project was approved by the Research Ethics Committees of the Universidade Federal do Rio Grande do Sul, Hospital de Clínicas de Porto Alegre, and the Universidade Estadual do Sudoeste da Bahia (Resolutions 18208/2010, 100565/2011, and 212/2010, respectively). All subjects signed an informed consent form approved by the above-mentioned ethics committees, according to the Helsinki Declaration.

2.2. MI measurements

The CANDELA protocol involved the acquisition of data on physical appearance, including a quantitative measure of skin color using the Derma spectrometer (DSM II ColorMeter, CyberDerm Inc., USA), which provides the MI, where in humans, values ranging from 20 to 100 reflect fairer to darker skin tones. MI was measured on the proximal medial portion of both arms to generate a single average value. This location was chosen because it captures the constitutive skin pigmentation (the basal quantity of melanin), with potentially little effect of sun exposure [5, 34].

2.3. Genetic analysis

DNA was extracted from total blood samples using the salting out method [35] or with the QIAamp® 96 DNA blood Kit (Qiagen), and genotyping procedures were performed with Taqman™ (KBioscience: www.kbioscience.co.uk). Eighteen

SNPs were investigated: rs1524668 (*ADAM17*), rs4785763 (*AFG3L1*, a pseudogene), rs6058017 (*ASIP*), rs1129038 (*HERC2*), rs1805009, rs1805008, and rs1805007 (*MC1R*), rs1800407, rs1800401, and rs1800414 (*OCA2*), rs1426654 (*SLC24A5*, also identified as *NCKX5*), rs6867641, rs26722, and rs16891982 (*SLC45A2*, also identified as *MATP*), rs3750965 and rs3829241 (*TPCN2*), and rs1042602 and rs1126809 (located at the *TYR* gene). These polymorphisms were chosen because previous studies have linked them either directly or indirectly with the human pigmentation pathway (Table S1).

2.4. Statistical analyses

Allele frequencies were estimated by gene counting. Agreement of genotype frequencies with Hardy-Weinberg equilibrium (HWE) was performed using the "Utility programs for analysis of genetic linkage" [36]. Three groups were considered: (a) total sample, (b) *Gaucho* sample only, and (c) *Baiano* sample only. Allele and genotype frequency differences were compared between *Gaucho* and *Baiano* samples with the Pearson χ^2 or Fisher's exact tests using the PEPI program, version 4.0 [37], or SPSS, version 17.0. The latter software was also used to conduct the analysis of variance (ANOVA) or the non-parametric Kruskal-Wallis test for comparing MI values across genotypes. The independent samples Student's *t*-test, or the non-parametric U value of the Wilcoxon-Mann-Whitney test, was used to compare MI and age means between *Gaucho* and *Baiano* samples. The Kolmogorov-Smirnov and Levene's tests were used to verify the normality of quantitative measures (MI and age). A p value of < 0.01 was considered statistically significant (SPSS, version 17.0).

The SPSS software was also used to perform regression analysis. Associations between common SNPs and MI were evaluated using a simple linear regression model, minimizing the sum of square differences between the observed (dependent) and predicted variables. Genotypes were re-coded as dummy variables based on the following criteria: (1,0), (0,1), and (0,0) for *AA*, *Aa*, and *aa*, respectively. Note that the inclusion of dummy variables in a linear regression model is equivalent to estimating the restrictive and non-restrictive models of linear regression using the usual least squares method. To avoid perfect multiple collinearity we used k-1 dummy variables. The R-value, or Pearson's correlation statistic, was obtained from

the ordinary least squares linear model. Coefficients of determination (R^2) were also calculated, whereas the *F* statistics tested the likelihood ratio of the explained variance based on the model of the residual sum of squares.

The STRUCTURE software [38] was used to verify if the markers studied have a power of differentiate between African and European gene pools, assuming that the Brazilian population is mainly composed by the intermixture of these two geographic human groups, besides Native Americans. Unfortunately, satisfying information for polymorphisms studied here is not available for Native Americans, so this group was not compared. According to Pritchard et al. [38], STRUCTURE implements a model-based clustering method for inferring population structure using genotype data consisting of unlinked markers. It assigns individuals to populations, identifying migrants and admixed individuals. Genotypes of t6wo reference ancestral groups (European and African) were available at The Human Genome Diversity Project (HGDP, http://hagsc.org/hgdp/files.html). STRUCTURE was run with 100,000 length of burnin period and 10,000 MCMC reps after burning, with k=2, k=3, and k=4.

3. Results

3.1 Management of quantitative data and gene frequency analyses

Age and MI absolute numbers did not present a Gaussian distribution; the numbers were therefore adjusted to normalise the data using 1/MI and 1/age transformations. The average age for the total sample (N = 503) was 25.19 (range 18–87) years, with 169 (33.6%) males and 334 (66.4%) females. The *Gaucho* average age was 26.19 (range 18–87) years, and the average *Baiano* age was 23.08 (18–62) years, which differed significantly ($p < 0.01$). The mean MI of the sample as a whole was 32.92 (range 22.04–76.40), while *Gaucho* and *Baiano* showed average MI values of 31.53 (22.04–76.40) and 36.75 (23.72–65.01), respectively. These results demonstrated that the *Baiano* group presents a darker skin color ($p < 0.01$) compared with the *Gaucho* group.

Within the *Gaucho* population, genotype frequencies were in accordance with those expected by HWE, except for *SLC24A5*rs1426654 ($\chi^2 = 11.346$, $p < 0.01$), which showed a lower number of heterozygotes than expected. For the *Baiano*

population, only the genotype frequencies of the SNP *SLC45A2*rs16891982 were not in HWE (χ^2 = 6.911, p < 0.01), which was also due to a lower number of heterozygotes than expected. Allele and genotype frequencies were statistically different between the 2 populations for *ASIP*rs6058017, *HERC2*rs1129038, *OCA2*rs1800401, *SLC24A5*rs1426654, and *SLC45A2*rs16891982, whereas the SNPs *MC1R*rs1805007, *TPCN2*rs3829241, *TYR*rs1042602, and *TYR*rs1126809 only differed in allele frequencies (Table S2).

3.2 Association between MI and genotypes

Three independent analyses were performed to evaluate the association between MI and genotype: considering the total sample (N = 503), the *Gaucho* sample only (N = 354), and the *Baiano* sample only (N = 149). Levene's test was applied to verify the variance homogeneity of the transformed data, along with ANOVA. For those showing heterogeneity, we performed the non-parametric Kruskal-Wallis test using untransformed MI values. In the total sample (Table 1), 8 polymorphisms showed a statistically significant association with MI (*ASIP*rs6058017, *HERC2*rs1129038, *OCA2*rs1800407, *OCA2*rs1800401, *SLC24A5*rs1426654, *SLC45A2*rs16891982, *TYR*rs1042602, and *TYR*rs1126809), at a 99% probability level. In the stratified analysis (Table 1), 4 SNPs (*HERC2*rs1129038, *MC1R*rs1805009, *SLC24A5*rs1426654, and *SLC45A2*rs16891982) were significantly associated with MI in the *Gaucho* sample, and 6 SNPs (*HERC2*rs1129038, *OCA2*rs1800407, *SLC24A5*rs1426654, *SLC45A2*rs16891982, *TYR*rs1042602, and *TYR*rs1126809) were associated with MI in the *Baiano* sample.

Therefore, 3 SNPs (*HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982) showed a significant association in all 3 analyses, with the heterozygous genotypes showing an intermediate MI value (Table 2, Figure 1). *SLC24A5*rs1426654 presented the highest MI differences between genotypes (*Gaucho*: GG = 46.49, GA = 34.29, AA = 30.80; *Baiano*: GG = 46.74, GA = 38.37, AA = 34.38, Table 2). The SNPs S*LC45A2*rs16891982 and *SLC24A5*rs1426654 were previously associated with skin color in European and/or Asian populations [6, 17, 18, 39, 40], and our results indicate a probable additive effect of alleles in these groups with distinct admixture histories. This effect was also observed by Stokowski et al. [41]. On the other hand, *HERC2*rs1129038 was previously associated only with

eye and hair color [42, 43]. This polymorphism is in high linkage disequilibrium with *HERC2*rs12913832 [43, 44], which, in turn, is associated with eye, hair, and skin color [40, 43]. Our study shows, for the first time, that *HERC2*rs1129038 also likely plays a role in skin color, at least in some populations, such as those investigated here. Population-specific associations were also found. For example, the *TYR*rs1042602/*TYR*rs1126809 and *MC1R*rs1805009 polymorphisms showed associations with skin color only in the *Baiano* or *Gaucho* groups, respectively (Table 1).

Interestingly, some studies have suggested that the C allele of *SLC45A2*rs16891982 was associated with darker skin and/or brown eyes, whereas in other studies, this allele was associated with lighter skin and/or blue eyes (Table S1). This discrepancy may be due to the fact that C/G differences are easily prone to reading mistakes. Regardless, our data indicated that the C allele (Phe) was related to higher MI values, thus corroborating the propositions of Norton et al. [6], Spichenok et al. [17], and Walsh et al., [18]. With respect to other markers, the alleles associated with darker skin pigmentation were *ASIP*rs6058017 G, *HERC2*rs1129038 G, *MC1R*rs1805009 G (Asp), *OCA2*rs1800407 G (Arg), *OCA2*rs1800401 T (Trp), *SLC24A5*rs1426654 G (Ala), *TYR*rs1042602 C (Ser), and *TYR*rs1126809 G (Arg) (Table S3).

The results of the regression analysis accorded with most of the results obtained using ANOVA or the Kruskal-Wallis test. This analysis showed that the *SLC24A5*rs1426654 and *SLC45A2*rs16891982 SNPs together explained 57%, 55%, and 64% of the total variation in MI values found in the *Gaucho, Baiano* and Total samples, respectively. The individual effects of all SNPs tested are shown in Table S4.

We also performed analysis to verify if 12 of the markers studied here are good ancestry predictors (*ADAM17*rs1524668, *AFG3L1*rs4785763, *MC1R*rs1805007, *OCA2*rs1800407, *OCA2*rs1800414, *SLC24A5*rs1426654, *SLC45A2*rs26722, *SLC45A2*rs16891982, *TPCN2*rs3750965, *TPCN2*rs3829241, *TYR*rs1042602, and *TYR*rs1126809) using the STRUCTURE software [38]. They were chosen because only for them there are data available for at least two putative Brazilian major parental stocks (European; n = 174 and African; n = 176). We performed the analysis with k = 4, k = 3, and k = 2. The best result, considering the mean value of likelihood (ln) provided by STRUCTURE using these 12 polymorphisms was with k = 2 (ln = -

7077.2). A clear African and non-African differentiation was observed with the SNP set (Figure 2).

4. Discussion

Knowledge about the connection between genetic variations and complex traits has improved markedly in the last few years. Nevertheless, current knowledge for genetic predictors of skin color remains limited [45], particularly with respect to populations outside of the European-North American axis. Our intention was to help fill this gap by demonstrating the effect of pigmentation genetic variants in 2 admixed Brazilian populations.

Our results showed that 3 polymorphisms (*HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982) seem to behave relatively consistently as skin color predictors, while the others seem to be population-specific, indicating that these SNPs should be avoided for general inferences in forensic DNA phenotyping. Note that differences in pigmentation are due to the amount and type of melanin synthesised in melanocytes, and to the shape and distribution of the melanosomes [3]. Complex gene and protein networks are related to these processes, and therefore, it is likely that their architectures differ, at least in part, across diverse human populations. Our findings have important forensic implications, since some authors have suggested that eye and hair color predictions can be developed independently of the biogeographic ancestry of the population investigated [17, 18, 46, 47]. Results of the present study indicated that for skin color, this assumption appears to be only partially true.

Despite the fact that we observed an intermediate effect of heterozygote genotypes in the majority of the significant SNPs, it is noteworthy that heterozygotes for some polymorphisms had MI means similar to those of one of the homozygotes. The description of specific effect of each genotype is extremely important to the knowledge of polymorphism effect. For instance, *OCA2*rs1800401 (total sample – data not shown), *SLC45A2*rs16891982, and *HERC2*rs1129038 (Table 2), suggest a completely dominant relationship between alleles, which was as also observed by Cook et al. [40] in a study of the expression of 2 of these genes across different genotype profiles.

Twelve markers studied here were able to differentiate Africans from non-Africans (Figure 1). We also observed that the admixture profile was similar between *Gaucho* and Europeans (Figure 1), and the *Baiano* sample had a slightly higher admixture level, compared with *Gaucho* group. This data corroborates demographic history reported for these populations. As early mentioned, our analysis also demonstrated that the *Baiano* group presents a darker skin color ($p < 0.01$) compared with the *Gaucho* group. The small differences between both sample could explain the different association of some markers studied here (for example: *MC1R*rs1805009, *OCA2*rs1800407, and *TYR*rs1042602; please see Table 1).

The present study is the first to investigate the relationship of the *ADAM17*rs1524668, *AFG3L1*rs4785763, *HERC2*rs1129038, *OCA2*rs1800407, and *TPCN2*rs3750965 polymorphisms with normal skin color variation. All of the other polymorphisms have already been associated with skin color in some populations around the world (Table S1), but had not been investigated in Brazilian samples until now.

5. Conclusion

Our results revealed that although 8 genetic markers were associated with skin color, important differences between the 2 admixed populations were found. Therefore, our association study demonstrates that genotype-phenotype relationships found in one population cannot be uncritically extrapolated to other populations with respect to skin color.

Acknowledgments

We are grateful to Prof. Sídia M. Callegari-Jacques for technical help and statistical support. We also thank the CANDELA Project Executive Committee members, and their volunteers in particular, for making this study possible.

References

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

[2] Hida T, Wakamatsu K, Sviderskaya EV, Donkin AJ, Montoliu L, Lamoreux ML, Yu B, Millhauser GL, Ito S, Barsh GS, Jimbow K, Bennett DC. (2009) Agouti protein, mahogunin, and attractin in pheomelanogenesis and melanoblast-like alteration of melanocytes: a cAMP-independent pathway. Pigment Cell Melanoma Res 22: 623–634.

[3] Parra EJ. (2007) Human pigmentation variation: evolution, genetic basis, and implications for public health. Am J Phys Anthropol 45: 85–105.

[4] Cheli Y, Luciani F, Khaled M, Beuret L, Bille K, Gounon P, Ortonne JP, Bertolotto C, Ballotti R. (2009) {alpha}MSH and Cyclic AMP elevating agents control melanosome pH through a protein kinase A-independent mechanism. J Biol Chem 284: 18699–18706.

[5] Quillen EE, Bauchet M, Bigham AW, Delgado-Burbano ME, Faust FX, Klimentidis YC, Mao X, Stoneking M, Shriver MD. (2012) OPRM1 and EGFR contribute to skin pigmentation differences between Indigenous Americans and Europeans. Hum Genet 131: 1073–1080.

[6] Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD. (2007) Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol 24: 710–722.

[7] Cerqueira CCS, Amorim CEG, Salzano FM, Bortolini MC. (2011) Human pigmentation genes: forensic perspectives, general aspects and evolution. In: Yacine N, Fellag R, editors. Forensic Science. New York: Nova Science Publishers Inc. pp. 85–106.

[8] Cerqueira CCS, Paixão-Côrtes VR, Zambra FMB, Salzano FM, Hünemeier T, Bortolini MC. (2012) Predicting *Homo* pigmentation phenotype through genomic data: from Neanderthal to James Watson. Am J Hum Biol 24: 705–709.

[9] McEvoy B, Beleza S, Shriver MD. (2006) The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. Hum Mol Genet 2: R176–R181.

[10] Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ. (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4: e1000074.

[11] Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, Hu FB, Thomas G, Hoover RN, Chanock S, Hunter DJ, Han J. (2009) Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol 129: 2250–2257.

[12] Gerstenblith MR, Shi J, Landi MT. (2010) Genome-wide association studies of pigmentation and skin cancer: a review and meta-analysis. Pigment Cell Melanoma Res 23: 587–606.

[13] Jablonski NG, Chaplin G. (2010) Human skin pigmentation as an adaptation to UV radiation. Proc Natl Acad Sci USA 107: 8962–8968.

[14] Jablonski JG, Chaplin G. (2012) Human skin pigmentation, migration and disease susceptibility. Phil Trans R Soc B 367: 785–792.

[15] Quillen EE, Shriver MD. (2011) Unpacking human evolution to find the genetic determinants of human skin pigmentation. J Invest Dermatol 131: E5–E7.

[16] Kayser M, Knijff P. (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nature 12: 179–192.

[17] Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, Caragine T, Prinz M, Wurmbach E. (2011) Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Sci Int Genet 5: 472–478.

[18] Walsh S, Liu F, Ballantyne KN, Van Oven M, Lao O, Kayser M. (2011) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5: 170–180.

[19] Paik SH, Kim HJ, Son HY, Lee S, Im SW, Ju YS, Yeon JH, Jo SJ, Eun HC, Seo JS, Kwon OS, Kim JI. (2012) Gene mapping study for constitutive skin color in an isolated Mongolian population. Exp Mol Med 44: 241–249.

[20] Fernandez LP, Milne RL, Pita G, Avilés JA, Lázaro P, Benítez J, Ribas G. (2008) SLC45A2: a novel malignant melanoma-associated gene. Hum Mutat 29: 1161–1167.

[21] Guedj M, Bourillon A, Combadières C, Rodero M, Dieudé P, Descamps V, Dupin N, Wolkenstein P, Aegerter P, Lebbe C, Basset-Seguin N, Prum B, Saiag P, Grandchamp B, Soufir N, MelanCohort Investigators. (2008) Variants of the MATP/SLC45A2 gene are protective for melanoma in the French population. Hum Mutat 29: 1154–1160.

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

[23] Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E, Shriver MD, Parra EJ, Rocha J. (2012) The timing of pigmentation lightening in Europeans. Mol Biol Evol 30: 24–35.

[24] Jablonski NG. (2012) The evolution of human skin colouration and its relevance to health in the modern world. J R Coll Physicians Edinb 42: 58-63.

[25] Sturm RA, Duffy DL. (2012) Human pigmentation genes under environmental selection. Genome Biol 13: 248–262.

[26] Salzano FM, Bortolini MC. (2002) The Evolution and Genetics of Latin American Populations. Cambridge: Cambridge University Press. 528 p.

[27] Guerreiro-Júnior V, Bisso-Machado R, Marrero A, Hünemeier T, Salzano FM, Bortolini MC. (2009) Genetic signatures of parental contribution in black and white populations in Brazil. Genet Mol Biol 32: 1–11.

[28] Neitzke-Montinelli V, Urmenyi TP, Rondinelli E, Cabello PH, Silva R, Moura-Neto RS. (2012) Polymorphisms upstream of the melanocortin-1 receptor coding region are associated with human pigmentation variation in a Brazilian population. Am J Hum Biol 24: 853–855.

[29] Rodrigues N. (2008) Os Africanos no Brasil. São Paulo: Editora Madras. 254 p.

[30] Bortolini MC, Silva WAJ, Guerra DC, Remonatto G, Mirandola R, Hutz MH, Weimer TA, Silva MC, Zago MA, Salzano FM. (1999) African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. Am J Hum Biol 11: 551–563.

[31] Priore MD, Venâncio RP. (2004) Ancestrais: uma Introdução à História da África Atlântica. Rio de Janeiro: Editora Campus. 208 p.

[32] Hünemeier T, Carvalho C, Marrero AR, Salzano FM, Pena SD, Bortolini MC. (2007) Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. Am J Phys Anthropol 133: 854–867.

[33] Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, Bortolini MC. (2007) Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. Am J Phys Anthropol 132: 301–310.

[34] Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA. (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112: 387–399.

[35] Lahiri DK, Nurnberger JI. (1991) A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleic Acids Res 19: 5444.

[36] Ott J. (1988) Program HWE ©, Utility Programs for Analysis of Genetic Linkage. Available: http://www.jurgott.org/linkage/ListSoftware.html. Accessed 12 August 2013.

[37] Abramson JH, Gahlinger PM. (2001) Computer programs for epidemiologists: PEPI. Version 4.0. Salt Lake City. UT: Sagebrush Press.

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

[39] Graf J, Hodgson R, Van Daal A. (2005) Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mutat 25: 278–284.

[40] Cook AL, Chen W, Thurber AE, Smit DJ, Smith AG, Bladen TG, Brown DL, Duffy DL, Pastorino L, Bianchi-Scarra G, Leonard JH, Stow JL, Sturm RA. (2009) Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. J Invest Dermatol 129: 392–405.

[41] Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, Ginger RS, Green MR, Van der Ouderaa FJ, Cox DR. (2007) A genomewide association study of skin pigmentation in a south Asian population. Am J Hum Genet 81: 1119–1132.

[42] Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L. (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet 123: 177–187.

[43] Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, Martin NG, Montgomery GW. (2008) A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet 82: 424-431.

[44] Amos CI, Wang L, Lee JE, Gershenwald JE, Chen WV, Fang S, Kosoy R, Zhang M, Qureshi AA, Vattathil S, Schacherer CW, Gardner JM, Wang Y, Bishop DT, Barrett JH, GenoMEL Investigators, MacGregor S, Hayward NK, Martin NG, Duffy DL, Q-Mega Investigators, Mann GJ, Cust A, Hopper J, AMFS Investigators, Brown KM, Grimm EA, Xu Y, Han Y, Jing K, McHugh C, Laurie CC, Doheny KF, Pugh EW, Seldin MF, Han J, Wei Q. (2011) Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. Hum Mol Genet 20: 5012–5023.

[45] Liu F, Wen B, Kayser M. (2013) Colorful DNA polymorphisms in humans. Semin Cell Dev Biol S1084-9521 (13) 00049-9. [Epub ahead of print].

[46] Valenzuela RK, Henderson MS, Walsh MH, Garrison NA, Kelch JT, Cohen-Barak O, Erickson DT, Meaney FJ, Walsh JB, Cheng KC, Ito S, Wakamatsu K, Frudakis T, Thomas M, Brilliant MH. (2010) Predicting phenotype from genotype: normal pigmentation. J Forensic Sci 55: 315–22.

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

Figure 1. Genotypes for three polymorphisms (rs16891982, rs1129038, and rs1426654) in the total sample showing heterozygous genotypes with intermediate MI values (confidence interval of 95%).

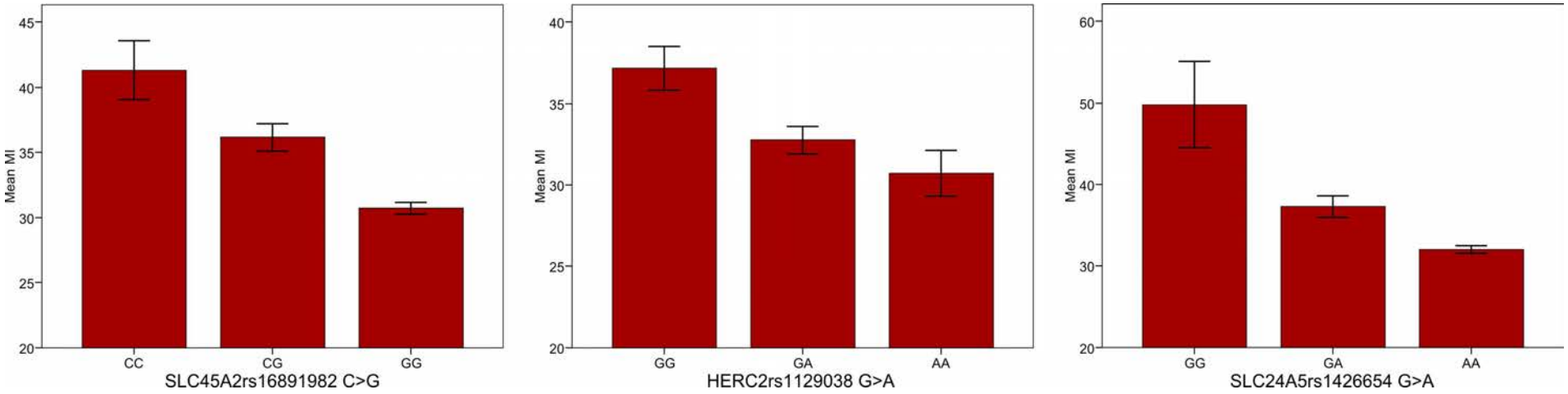


Figure 2. Analysis with the STRUCTURE software, with *k* = 2 using 12 of the 18 polymorphisms analyzed in this study. The black vertical lines split the groups analyzed.

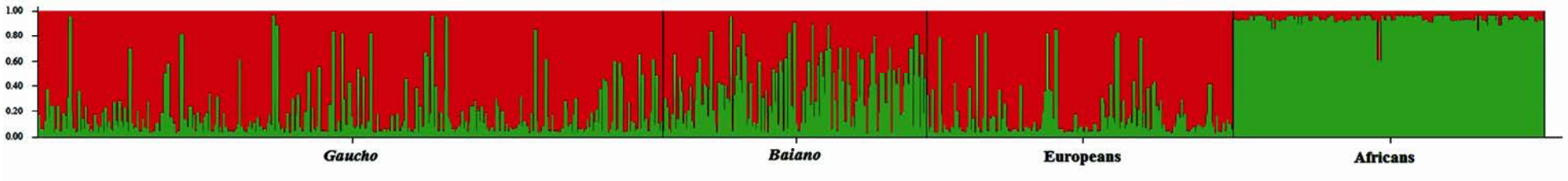


Table 1. Association of genotypes with Melanin Index (MI) in the 2 groups studied.

		Total sample (N = 503)		***Gaucho* (N = 354)**		***Baiano* (N = 149)**	
Gene	**SNP**	**Genotype x MI**		**Genotype x MI**		**Genotype x MI**	
		F[a] or X^{2b}	***p***	**F[a] or X^{2b}**	***p***	**F[a] or X^{2b}**	***p***
ADAM17	rs1524668	0.009[a]	0.991	0.907[a]	0.405	0.578[a]	0.562
AFG3L1	rs4785763	0.547[a]	0.579	0.464[a]	0.629	0.193[b]	0.911
ASIP	rs6058017	**12.475[b]**	**<0.01**	0.743[b]	0.696	1.194[a]	0.306
HERC2	rs1129038	**40.901[b]**	**<0.01**	**5.374[a]**	**<0.01**	**5.108[a]**	**<0.01**
MC1R	rs1805009	5.252[a]	0.022	**7.488[a]**	**<0.01**	0.014[b]	0.911
MC1R	rs1805008	0.024[a]	0.878	0.014[a]	0.905	N/A[c]	N/A[c]
MC1R	rs1805007	3.473[a]	0.063	0.952[a]	0.330	0.079[a]	0.779
OCA2	rs1800407	**8.589[a]**	**<0.01**	3.423[a]	0.065	**8.632[a]**	**<0.01**
OCA2	rs1800401	**4.626[a]**	**0.01**	1.509[a]	0.220	2.026[a]	0.136
OCA2	rs1800414	0.071[a]	0.790	0.495[a]	0.482	0.185[a]	0.667
SLC24A5	rs1426654	**95.188[b]**	**<0.01**	**51.902[a]**	**<0.01**	**16.562[a]**	**<0.01**
SLC45A2	rs6867641	1.377[a]	0.255	1.994	0.369	9.760[a]	0.089
SLC45A2	rs26722	1.088[a]	0.338	2.818[a]	0.061	0.122[a]	0.885
SLC45A2	rs16891982	**158.876 [b]**	**<0.01**	**73.769[b]**	**<0.01**	**38.285[a]**	**<0.01**
TPCN2	rs3750965	1.330[a]	0.265	0.844[a]	0.431	0.026[a]	0.975
TPCN2	rs3829241	2.078[a]	0.126	2.676[a]	0.070	0.142[a]	0.868
TYR	rs1042602	**15.346 [b]**	**<0.01**	3.079[b]	0.214	**7.224[a]**	**<0.01**
TYR	rs1126809	**9.335[a]**	**<0.01**	3.554[a]	0.030	**5.028[a]**	**<0.01**

[a]ANOVA with MI transformed to 1/MI to fit a symmetrical distribution. [b]Nonparametric Kruskal-Wallis test was used when variances were not homogenous even after transformation, as verified by Levene's test. [c]N/A: *MC1R*rs1805008 had a small sample number (N = 24) in the *Baiano* subgroup, being monomorphic for the CC genotype. For this reason, it was not possible to run the analysis for this SNP.

Table 2. MI values by genotype in 3 statistically significant polymorphisms.

	Total sample	***Gaucho***	***Baiano***
		***HERC2*rs1129038**	
	Average ranking (N)	MI (N)	MI (N)
GG	205.86[a] (157)	32.38[b] (65)	38.23[b] (92)
GA	152.15[a] (141)	31.06[b] (95)	34.27[b] (46)
AA	110.89[a] (44)	29.89[b] (41)	34.05[b] (3)
		***SLC24A5*rs1426654**	
	Average ranking (N)	MI (N)	MI (N)
GG	429.26[a] (23)	46.49[b] (9)	46.74[b] (14)
GA	332.22[a] (104)	34.29[b] (51)	38.37[b] (53)
AA	212.87[a] (368)	30.80[b] (290)	34.38[b] (78)
		***SLC45A2*rs16891982**	
	Average ranking (N)	Average ranking (N)	MI (N)
CC	377.14[a] (66)	269.40[a] (24)	40.89[b] (42)
CG	301.13[a] (164)	220.13[a] (110)	39.40[b] (54)
GG	169.66[a] (254)	137.46[a] (211)	31.18[b] (43)

[a]Kruskal-Wallis test results, the values are the mean rank; [b]MI values based on transformed 1/MI.

Table S1. Characteristics of the SNPs studied in this work[a].

Gene	SNP ID	Localization	Number of exons	Mutation	Probable allele function	Studied populations	Reference (s)
ADAM17[b]	rs1524668	2p25.1	19	-968C>A - 5'UTR	-	Admixed Brazilian populations	Present study
AFG3L1	rs4785763	16q24.3	pseudogene	A>C	A = red hair, freckles/ C = not-red hair, no freckles	Europeans	[1]
ASIP	rs6058017	20q11.22	4	8818G>A - 3'UTR	G = more melanin, no freckles/ A= less melanin, freckles	African-Americans from Washington, Europeans, East Asians	[1, 2, 3, 4]
HERC2	rs1129038	15q13.1	93	G>A - 3'UTR	A = blue eyes/ G = not-blue eyes	Europeans	[5, 6]
MC1R	rs1805009	16q24.3	1	Asp294His - G>C	G = not-red hair, not-fairer skin/ C= red hair, fairer skin	Europeans, Caucasians	[7, 8, 9]
	rs1805008			Arg160Trp - C>T	T = blonde or red hair, freckles, skin sensitivity to sun, fairer skin/ C = not-blonde or not-red hair, no freckles	Europeans, Caucasians	[1,8, 9, 10]
	rs1805007			Arg151Cys - C>T	T = blonde or red hair, freckles, skin sensitivity to sun, fairer skin/ C = not-blonde or not-red hair, no freckles	Europeans, Caucasians	[1, 8, 9, 10]
OCA2	rs1800407	15q13.1	24	Arg419Gln - G>A	G = not-green or brown or black eyes/ A = green or blue eyes	Caucasians of Pennsylvania, Europeans	[11, 12]
	rs1800401			Arg305Trp - C>T	C = blue eyes, lighter pigmentation/ T = brown/black eyes and darker pigmentation	Caucasians of Pennsylvania	[11]
	rs1800414			His615Arg - A>G	A = darker skin/ G = fairer skin	East Asian ancestry living in Canada	[13]
SLC24A5 (NCKX5)	rs1426654	15q21.1	9	Ala111Thr - G>A	G = more melanin/ A = less melanin	Europeans, Mixed populations, African–Americans, Asians	[4, 14, 15]
SLC45A2 (MATP)	rs6867641	5p13.3	7	-1169C>T - 5'UTR	C = fairer skin/ T = darker skin	Caucasians, Asians, African Americans, Australian Aborigines, Spanish Basques	[16]
	rs26722			Glu272Lys - G>A	G = fairer skin, lighter brown hair, not-brown eyes/ A = darker skin, hair, and eye color	Caucasians, Asians, African-Americans, Australian Aborigines	[17]
	rs16891982			Phe374Leu - C>G	C = darker skin, brown or black eyes/ G = blue eyes, lighter or medium skin, darker skin, hair, and eye color	Caucasians, Asians, African-Americans, Australian Aborigines, Europeans, Mixed populations	[4, 12, 14, 15, 17]
TPCN2	rs3750965	11q13.2	25	Lys376Arg - A>G	A = blonde hair / G = brown hair	Europeans	[1]
	rs3829241			Gly734Glu - G>A	A= blond hair, skin sensitivity to sun/ G= brown hair, darker skin	Europeans	[1]

TYR	rs1042602	11q14.3	5	Ser192Tyr - C>A	C = skin sensitivity to sun, freckles, brown hair/ A = no freckles, blonde hair	Europeans	[1, 10]
	rs1126809			Arg402Gln - G>A	G= not-blue eyes, brown hair/ A=blue eyes, blonde hair, skin sensitivity to sun	Europeans	[1]

[a] Key of abbreviations: A = Adenine, *ADAM17* = a disintegrin and metallopeptidase domain 17, *AFG3L1* = ATPase family gene 3-like 1 (pseudogene), Ala = alanine, *ASIP* = agouti signalling protein, Arg = arginine, Asp = aspartic acid, C = Cytosine, G = Guanine, Gln = glutamine, Glu = glutamic acid, Gly = glycine, *HERC2* = HECT and RLD domain containing E3 ubiquitin protein ligase 2, His = histidine, Leu = leucine, Lys = lysine, *MATP* = membrane associated transporter protein, *MC1R* = melanocortin 1 receptor, *NCKX5*= Na/Ca/K exchanger 5, *OCA2* = oculocutaneous albinism II, Phe = phenylalanine, Ser = serine, *SLC24A5* = solute carrier family 24, member 5, *SLC45A2* = solute carrier family 45 member 2, T= Thymine, Thr = threonine, *TPCN2* = two pore segment channel 2, Trp = tryptophan, *TYR* = tyrosinase (gene), Tyr = tyrosine (amino acid), UTR = untranslated region.

[b] *ADAM17* was mentioned by Norton et al. [4] as an 'a priori pigmentation candidate gene', with a positive selection signal in East-Asians. We selected the *ADAM17*rs1524668 SNP because the derived allele is in high frequency in Amerindians (HGDP-CEPH, http://www.cephb.fr/en/hgdp/diversity.php/).

Table S2. Genotype and allele frequencies for 18 SNPs in the *Gaucho* (N = 354) and *Baiano* (N = 149) samples.

	rs1042602		rs1126809		rs6867641		rs26722		rs16891982		rs1800407	
	1 = C/2 = A		1 = G/2 =A		1 = C/2 = T		1 = C/2 = T		1 = C/2 = G		1 = G/2 = A	
Genotypes/alleles	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano [a]	Gaucho	Baiano
11	0.371	0.483	0.667	0.752	0.484	0.000	0.826	0.815	**0.069**	**0.302**	0.881	0.875
12	0.457	0.392	0.287	0.234	0.410	0.750	0.171	0.178	**0.319**	**0.388**	0.119	0.125
22	0.172	0.125	0.046	0.014	0.106	0.250	0.003	0.007	**0.612**	**0.310**	0.000	0.000
1	**0.599**	**0.678**	**0.815**	**0.869**	0.689	0.375	0.912	0.904	**0.229**	**0.496**	0.941	0.938
2	**0.401**	**0.322**	**0.185**	**0.131**	0.311	0.625	0.088	0.096	**0.771**	**0.504**	0.059	0.062
	rs1800401		rs1800414		rs1129038		rs6058017		rs3750965		rs3829241	
	1 = C/2 = T		1 = A/2 = G		1 = G/2 = A		1 = G/2 = A		1 = A/2 = G		1 = G/2 = A	
Genotypes/alleles	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano
11	**0.883**	**0.799**	0.994	0.993	**0.323**	**0.653**	**0.012**	**0.095**	0.437	0.514	0.487	0.587
12	**0.117**	**0.188**	0.006	0.070	**0.473**	**0.326**	**0.226**	**0.361**	0.449	0.363	0.399	0.371
22	**0.000**	**0.013**	0.000	0.000	**0.204**	**0.021**	**0.762**	**0.544**	0.114	0.123	0.114	0.042
1	**0.942**	**0.893**	0.997	0.997	**0.560**	**0.816**	**0.125**	**0.276**	0.662	0.695	**0.686**	**0.773**
2	**0.058**	**0.107**	0.003	0.003	**0.440**	**0.184**	**0.875**	**0.724**	0.338	0.305	**0.314**	**0.227**
	rs1426654		rs1524668		rs1805009		rs1805008		rs1805007		rs4785763	
	1 = G/2 = A		1 = G/2 = T		1 = G/2 = C		1 = C/2 = T		1 = C/2 = T		1 = A/2 = C	
Genotypes/alleles	Gaucho [a]	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano	Gaucho	Baiano
11	**0.026**	**0.096**	0.405	0.432	0.971	0.986	0.914	1.000	0.951	0.993	0.045	0.056
12	**0.146**	**0.366**	0.462	0.480	0.029	0.014	0.086	0.000	0.049	0.007	0.368	0.347
22	**0.828**	**0.538**	0.133	0.088	0.000	0.000	0.000	0.000	0.000	0.000	0.587	0.597
1	**0.099**	**0.279**	0.636	0.672	0.986	0.993	0.957	1.000	**0.973**	**0.997**	0.228	0.229
2	**0.901**	**0.721**	0.364	0.328	0.014	0.007	0.043	0.000	**0.027**	**0.003**	0.772	0.771

[a]The rs values deviating from Hardy-Weinberg Equilibrium, indicated in each subgroup;
Allele and genotype frequencies with statistically significant differences ($p < 0.01$) between *Gaucho* and *Baiano* samples are in bold.

Table S3. Summary of the associations found in this study.

Genes	Polymorphisms	Total sample	*Gaucho*	*Baiano*
ADAM17	rs1524668 G>T	N/A	N/A	N/A
AFG3L1	rs4785763 A>C	N/A	N/A	N/A
ASIP	rs6058017 G>A	GG = ↑ MI/ AA = ↓ MI	N/A	N/A
HERC2	rs1129038 G>A	GG = ↑ MI/ AA = ↓ MI	GG = ↑ MI/ AA = ↓ MI	GG = ↑ MI/ AA = ↓ MI
MC1R	rs1805009 G>C	N/A	GG = ↑ MI[a]	N/A
MC1R	rs1805008 C>T	N/A	N/A	N/A
MC1R	rs1805007 C>T	N/A	N/A	N/A
OCA2	rs1800407 G>A	GG = ↑ MI[a]	N/A	GG = ↑ MI[a]
OCA2	rs1800401 C>T	TT = ↑ MI/ CC = ↓ MI	N/A	N/A
OCA2	rs1800414 A>G	N/A	N/A	N/A
SLC24A5	rs1426654 G>A	GG = ↑ MI/ AA = ↓ MI	GG = ↑ MI/ AA = ↓ MI	GG = ↑ MI/ AA = ↓ MI
SLC45A2	rs6867641 C>T	N/A	N/A	N/A
SLC45A2	rs26722 G>A	N/A	N/A	N/A
SLC45A2	rs16891982 C>G	CC = ↑ MI/ GG = ↓ MI	CC = ↑ MI/ GG = ↓ MI	CC = ↑ MI/ GG = ↓ MI
TPCN2	rs3750965 A>G	N/A	N/A	N/A
TPCN2	rs3829241 G>A	N/A	N/A	N/A
TYR	rs1042602 C>A	CC = ↑ MI/ AA = ↓ MI	N/A	CC = ↑ MI/ AA = ↓ MI
TYR	rs1126809 G>A	GG = ↑ MI/ AA = ↓ MI	N/A	GG = ↑ MI/ AA = ↓ MI

N/A = No Association.

[a]We did not find the other homozygote in this sample.

Table S4. Regression analysis between SNPs and MI for the total sample, *Gaucho* subgroup, and *Baiano* subgroup.

Total Sample					
SNP	**R**	**R-squared**	**Corrected**	**F**	***p*-value (model)[a]**
rs1524668	0.015505	0.00024039	0.00383195	0.05903044	0.942685
rs4785763	0.071004	0.00504161	0.00087861	121.105.068	0.298796
rs6058017	0.22425953	0.05029234	0.04633522	12.709.343	**0.000004**
rs1129038	0.348074	0.1211553	0.11597037	233.668.388	**0.000000**
rs1805009	0.054139	0.00293102	0.00090446	144.630.208	0.229701
rs1805008	0.019741	0.0003897	0.00566854	0.06432586	0.800101
rs1805007	0.07885	0.00621732	0.00421373	310.308.599	0.07876
rs1800407	0.107682	0.01159552	0.00956176	57.015.344	0.017331
rs1800401	0.156441	0.02447383	0.02048395	613.397.347	**0.002338**
rs1800414	0.00151266	0.0000022881	-0.0020261	0.00112806	0.97322
rs1426654	0.58586759	0.34324083	0.34057108	128.566.525	**0.000000**
rs6867641	0.14798556	0.02189973	0.00982441	181.359.507	0.166362
rs26722	0.02887526	0.00083378	0.00325279	0.20402953	0.815508
rs16891982	0.54340697	0.29529113	0.29236095	100.775.683	**0.000000**
rs3750965	0.08411019	0.00707452	0.00298841	173.135.771	0.178134
rs3829241	0.09502396	0.00902955	0.00490909	219.139.502	0.112875
rs1042602	0.20153652	0.04061697	0.03677944	105.841.378	**0.000031**
rs1126809	0.17752206	0.03151408	0.02753673	792.337.681	**0.000411**
Gaucho					
SNP	**R**	**R-squared**	**Corrected**	**F**	***p*-value (model)[a]**
rs1524668	0.06053178	0.0036641	-0.00214544	0.63070339	0.532833
rs4785763	0.06390313	0.00408361	-0.00187996	0.68475912	0.504918
rs6058017	0.202474304	0.04099584	0.03523606	711.760.003	**0.000940**
rs1129038	0.21974732	0.04828889	0.03867564	502.316.253	**0.007447**
rs1805009	0.11429611	0.0130636	0.01020292	456.659.849	0.033304
rs1805008	0.01868701	0.0003492	-0.00585981	0.05624155	0.81284
rs1805007	0.04735991	0.00224296	-0.00062416	0.782305	0.377048
rs1800407	0.09386853	0.0088113	0.00591309	304.025.319	0.082122

rs1800401	0.07396395	0.00547067	0.00255416	187.575.852	0.171718
rs1800414	0.024972122	0.00062361	-0.00225644	0.21652661	0.641991
rs1426654	0.585899952	0.34327875	0.33949362	906.912.395	**0.000000**
rs6867641	0.148378244	0.0220161	0.00963656	177.842.619	0.172268
rs26722	0.096792774	0.00936884	0.00359257	162.195.205	0.199024
rs16891982	0.476287386	0.22684967	0.22232833	501.730.298	**0.000000**
rs3750965	0.077957342	0.00607735	0.00023074	103.946.622	0.354764
rs3829241	0.135555829	0.01837538	0.01256695	316.357.154	0.043529
rs1042602	0.157303218	0.0247443	0.01918729	445.280.666	0.012311
rs1126809	0.118407897	0.01402043	0.00825447	243.158.541	0.089416

Baiano

SNP	R	R-squared	Corrected	F	*p*-value (model)[a]
rs1524668	0.08497432	0.00722064	-0.00647287	0.52730351	0.591322
rs4785763	0.10137915	0.01027773	-0.00376088	0.73210441	0.482715
rs6058017	0.144564188	0.0208988	0.00730018	153.683.187	0.218568
rs1129038	0.26214544	0.06872023	0.05522342	509.159.144	**0.007354**
rs1805009	0.11046328	0.01220214	0.00538974	179.116.571	0.182879
rs1805008	0.04735991	0.00224296	-0.00062416	0.782305	0.377048
rs1805007	0.03497655	0.00122336	-0.00561758	0.17882921	0.673004
rs1800407	0.16978261	0.02882613	0.02198688	421.480.765	0.041908
rs1800401	0.18106973	0.03278625	0.01953674	247.452.656	0.087728
rs1800414	0.044882538	0.00201444	-0.00491601	0.29066521	0.590627
rs1426654	0.508462386	0.258534	0.24809082	247.562.448	**0.000000**
rs6867641	0.921161704	0.84853888	0.77280833	112.047.093	0.078838
rs26722	0.07573842	0.00573631	-0.00816948	0.41251234	0.66277
rs16891982	0.538916029	0.29043049	0.27999564	27.832.753	**0.000000**
rs3750965	0.046660936	0.00217724	-0.01177832	0.15601255	0.855694
rs3829241	0.028203932	0.00079546	-0.01347889	0.05572665	0.945819
rs1042602	0.270925657	0.07340071	0.06070757	578.270.673	**0.003829**
rs1126809	0.246370821	0.06069858	0.04746898	458.808.981	0.011726

[a] Significance at $p < 0.01$ are indicated in bold.

Supporting Information References

[1] Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G, Pálsson S, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Vermeulen SH, Goldstein AM, Tucker MA, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K. (2008) Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 40: 835–837.

[2] Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR. (2002) A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet 70: 770–775.

[3] Bonilla C, Boxill LA, Donald SA, Williams T, Sylvester N, Parra EJ, Dios S, Norton HL, Shriver MD, Kittles RA. (2005) The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. Hum Genet 116: 402–406.

[4] Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD. (2007) Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol 24: 710–722.

[5] Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L. (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet 123: 177–187.

[6] Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, Martin NG, Montgomery GW. (2008) A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet 82: 424–431.

[7] Smith R, Healy E, Siddiqui S, Flanagan N, Steijlen PM, Rosdahl I, Jacques JP, Rogers S, Turner R, Jackson IJ, Birch-Machin MA, Rees JL. (1998) Melanocortin 1 receptor variants in an Irish population. J Invest Dermatol 111: 119–122.

[8] Han J, Kraft P, Colditz GA, Wong J, Hunter DJ. (2006) Melanocortin 1 receptor variants and skin cancer risk. Int J Cancer 119: 1976–1984.

[9] Latreille J, Ezzedine K, Elfakir A, Ambroisine L, Gardinier S, Galan P, Hercberg S, Gruber F, Rees J, Tschachler E, Guinot C. (2009) MC1R gene polymorphism affects skin color and phenotypic features related to sun sensitivity in a population of French adult women. Photochem Photobiol 85: 1451–1458.

[10] Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, Jakobsdottir M, Steinberg S, Pálsson S, Jonasson F, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K. (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39: 1443–1452.

[11] Rebbeck TR, Kanetsky PA, Walker AH, Holmes R, Halpern AC, Schuchter LM, Elder DE, Guerry D. (2002) P gene as an inherited biomarker of human eye color. Cancer Epid. Biomarkers Prev 11: 782–784.

[12] Walsh S, Liu F, Ballantyne KN, Van Oven M, Lao O, Kayser M. (2011) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5: 170–180.

[13] Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, Jin L, Parra EJ. (2010) Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: Further evidence of convergent evolution of skin pigmentation. PLoS Genet 6: e1000867.

[14] Cook AL, Chen W, Thurber AE, Smit DJ, Smith AG, Bladen TG, Brown DL, Duffy DL, Pastorino L, Bianchi-Scarra G, Leonard JH, Stow JL, Sturm RA. (2009) Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. J Invest Dermatol 129: 392–405.

[15] Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, Caragine T, Prinz M, Wurmbach E. (2011) Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Sci Int Genet 5: 472–478.

[16] Graf J, Voisey J, Hughes I, Van Daal A. (2007) Promoter polymorphisms in the MATP (SLC45A2) gene are associated with normal human skin color variation. Hum Mut 28: 710–717.

[17] Graf J, Hodgson R, Van Daal A. (2005) Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mut 25: 278–284.

3.3. CAPÍTULO III

“Novel associations of SNPs with eye and hair pigmentation in an admixed population**”**

(Manuscrito em preparação, versão I).

Novel associations of SNPs with eye and hair pigmentation in an admixed population

C.C.S. Cerqueira[1], T. Hünemeier[1], V. Ramallo[1], A.A.L. Barbosa[2], J.A. Gomez-Valdes[3], D. Longo[4], F. Rothhammer[5], G. Bedoya[6], S. Canizales-Quinteros[7,8], V. Acuña-lonzo[9], C. Gallo[10], G. Poletti[10], R. González-José[11], S. de Azevedo[11], F.M. Salzano[1], L. Schuler-Faccini[1,12], A. Ruiz-Linares[13], M.C. Bortolini[1]* for CANDELA (*Consortium for the Analysis of the Diversity and Evolution of Latin America*).

1 Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

2 Departamento de Ciências Biológicas, Universidade Estadual do Sudoeste da Bahia, Jequié, Brasil

3 Laboratorio de Antropología Física, Departamento de Anatomía, Facultad de Medicina, Universidad Nacional Autónoma de Mexico, Mexico City, Mexico

4 Instituto Federal Farroupilha, Farroupilha, Brazil

5 Instituto de Alta Investigación, Universidad de Tarapacá, Facultad de Medicina, Universidad de Chile and Centro de Investigaciones del Hombre en el Desierto, Arica, Chile

6 Universidad de Antioquia, Medellín, Colombia

7 Unidad de Genómica de Poblaciones Aplicada a la Salud, Facultad de Química, Universidad Nacional Autónoma de México, Mexico City, Mexico

8 Instituto Nacional de Medicina Genómica, Mexico City, Mexico

9 Molecular Genetics Laboratory, Escuela Nacional de Antropología e Historia, Mexico City, Mexico

10 Laboratorio de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú

11 Centro Nacional Patagónico - CENPAT, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Puerto Madryn, Argentina

12 INAGEMP – Instituto Nacional de Genética Médica Populacional, Porto Alegre, Brazil

13 Department of Genetics, Evolution and Environment and UCL Genetics Institute, University College London, London, UK

*Corresponding author: Tel.: +55 51 3308 9844. E-mail address: maria.bortolini@ufrgs.br

ABSTRACT

The first impressions about eye color indicated to be dealing with a Mendelian inheritance. This misconception was soon demystified by other studies. Currently, we know that traits like eye and hair color is influenced by many genes. Today, a handful of studies have been done trying to clarify the genetic patterns of human pigmentation. Unfortunately, these studies have been done predominantly in Europeans and/or Asians populations, not covering the long range spectrum of human normal variation. Admixed populations, like Brazilians, may therefore also be useful in the identification of specific genes involved in phenotypic variation (normal or pathological) of pigmentation. Here, we study the association between 18 SNPs with eye and hair color in an admixed population to help to clarify the genotype-phenotype relationship of these features, with the goal of bringing information that may be used in a wider reality, such as forensic genetics. We report also the linkage disequilibrium between some these markers. In an attempt to produce reliable data, we focus on performing high accuracy and high level of detail in color phenotype description. Our data indicated 3 markers (*HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982) highly associated with eye and hair color. These associations observed were very consistent between qualitative and quantitative measures of eye and hair phenotypes. Besides that, we observed a quantitative profile of RGB values between two phenotypic categories of eye color, corroborating literature data indicating a more predictability to forensic practice in those phenotypes. Despite the efforts of the scientific community in quantifying phenotypes, the use of qualitative features can bring less confusion in defining cutoff points of phenotypes for use in forensic DNA phenotyping.

Key-words: Brazillians, genetic patterns, human pigmentation, phenotypic categories, qualitative and quantitative measures.

1. Introduction

Since Davenport[1] and Hurst[2] published the first impressions about eye color inheritance, a handful of studies have been done trying to clarify the genetic patterns of iris color. The misconception of a Mendelian inheritance of eye color was soon demystified by other two studies.[3,4] Currently, we know that traits like eye and hair color is influenced by many genes.[5,6] Unfortunately, the majority of the pigmentation genetic studies in humans have been done in European and/or Asian populations.[7-11], not covering the long range spectrum of human normal variation. Furthermore, we know that these associations in a population cannot be easily transposed to another because of different patterns of linkage disequilibrium (LD), highlighting the importance of studies in specific populations.

The skin, eye and hair color are different between African, European and Native Americans, three of the parental groups that contributed to the formation of the Latin American population.[12] Despite of the major variability of eye and hair color to be geographically restricted in Europeans,[13] admixed population show some heterogeneity in these traits, due mainly to European inheritance. Brazilian populations, with their stories of relatively recent and different patterns of admixture[12,14-16] may therefore also be useful in the identification of specific genes involved in phenotypic variation (normal or pathological) of pigmentation. In this context, an initial analysis of individual polymorphisms involved in phenotypic traits within a population can certainly help to clarify the genotype-phenotype relationship of this features so that, subsequently, the data can also be used and possibly validated in macro populations, with the goal of bringing information that may be used in a wider reality.

It is believed that all of our externally visible characteristics (besides sex), including eye and hair color, are currently the most promising for DNA prediction.[17] The normal variation of human pigmentation was first investigated by medical scientists[18,19], anthropologists[20,21] and, more recently, by researchers of forensic genetics to predicting physical features by DNA[22,23], applying as an investigative tool. In this way, in recent years, numerous studies have identified genes responsible for such visible phenotypic characteristics[24-27], making real the promise to predict physical characteristics of an individual by DNA, initiating what was called Forensic DNA Phenotyping (FDP) or Intelligence by DNA.[22,28]. Recently, Walsh *et al.*[23] conducted an

important study for FDP technology, showing the ability to predict eye and hair color using 24 pigmentation genetics markers (HIrisplex system) in Europeans.

Here, we investigate the association of 18 SNPs (*ADAM17*rs1524668, *AFG3L1*rs4785763, *ASIP*rs6058017, *HERC2*rs1129038, *MC1R*rs1805007, *MC1R*rs1805008, *MC1R*rs1805009, *OCA2*rs1800401, *OCA2*rs1800407, *OCA2*rs1800414, *SLC24A2*rs1426654, *SLC45A2*rs6867641, *SLC45A2*rs26722, *SLC45A2*rs16891982, *TPCN2*rs3750965, *TPCN2*rs3829241, *TYR*rs1042602, and *TYR*rs1126809), including 6 SNPs used in HIrisplex system, with eye and hair color in an admixed Brazilian sample, comparing the consistence of the association among quantitative and qualitative measures, and also we report the LD between the markers analyzed.

2. Material and Methods

2.1 Recruitment of subjects of this study

All 563 Brazilian subjects of this work are also volunteers of the Consortium for the Analysis of the Diversity and Evolution of Latin America (CANDELA, *http://www.ucl.ac.uk/silva/candela*). This project involves academic researchers studying biological diversity of Latin Americans considering also their social contexts. More information about the sample used here can be found in Cerqueira *et al.*[29] Additional characteristics of the sample can be view in Supplementary Table 1. All subjects signed an informed consent approved by Ethic Committees of Universidade Federal do Rio Grande do Sul (UFRGS), Hospital de Clínicas de Porto Alegre (HCPA) and Universidade Estadual do Sudoeste da Bahia (UESB) (Resolutions 18208/2010, 100565/2011, and 212/2010, respectively).

2.2 Qualitative and quantitative measures of eye and hair color

During the sample collection of this project was filled a qualitative questionnaire about eyes and hair color of volunteers. This qualification was made and checking by at least, two CANDELA researchers (CCSC, VR and TH), avoiding personal impressions in the analysis. In case of disagreement between the two observers, a third opinion was required. This can to prevent biases in hetero-classification of some qualitative

characteristics, as observed by Bastos *et al.*[30] In summary, the eye color of the subjects were categorized into one of five categories, they are: Blue/grey, Green, Honey, Light Brown, and Dark Brown/Black; the hair color, in turn, were qualify into one of four categories, as follow: Red/reddish, Blond, Dark blond/Light Brown, and Dark Brown/Black. Similar classifications are commonly found in the literature.[13, 31-33]

An effort to quantify the eyes and hair color of subjects from the photos was also performed based on the protocol by Liu *et al.*[34] The photos of the volunteers were used to measure the RGB value in the region of the eyes and hair using the software AxioVision 4.8.2. The RGB values (from 0 to 255) were then transformed into Hue and Saturation values (from 0 to 1; HSV/HSB system), using standards formulas. In the following link it is possible to transform values of various systems of color measurements to each other through simple math, as well as to view the formulas used for these conversions: *http://www.easyrgb.com/index.php?X=MATH*. Genetic analyzes were performed using RGB and HS measures.

2.3. Genotyping and statistical analyses

The subjects were genotyped by TaqMan® for the 18 SNPs previously associated with human pigmentation in Europeans and/or Asians. Detailed information about these markers can be found in Cerqueira *et al.*[29] In the present work was utilized the same sample of Cerqueira *et al.*[29], that is, subjects from Rio Grande do Sul and Bahia states, plus 60 volunteers from the others Brazilian regions. To verify association between SNPs and phenotypes, we used the Fisher's exact test or ANOVA (for qualitative and quantitative data, respectively), and also Kruskal-Wallis in substitution to ANOVA, when necessary, using the SPSS, version 17.0. For all analysis a p value <0.001 was considered significant. The MLocus software[35] was used to verify LD between markers localized in the same or close genes.

3. Results

3.1 Association analysis with qualitative data

All genotypic and allele frequencies of markers analyzed are describe in Supplementary Table 2. Two SNPs (*SLC24A5*rs1426654 and *SLC45A2*rs16891982) are

not in Hard-Weinberg Equilibrium, explained by smaller heterozygous number than expected. In qualitative association analysis of the 18 markers with the phenotype of eye and hair color, we observed that 3/18 markers were highly associated with the eye color phenotype (*HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982), and we also verified that these same markers were significantly associated with hair color in our sample. The summary of these results are in Table 01 (more detailed information can be viewed on Supplementary Table 3 and 4).

3.2 Association analysis with quantitative data

In quantitative association analysis, we observed that the *HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982 also shown highly associated with both eye and hair color ($p<0.001$). This association was significant with R, G and B values, as well as with Hue measure (eye color) and Saturation measure (hair color) (Table 1). Due to quantity analyses done, we consider significance with $p<0.001$. As the *p* values was similar with the right and left eyes analyses, all *p* values showed in Table 1 to eye color are relative to right eye. In Supplementary Table 5 we present mean values to Red, Green, Blue, Hue, and Saturation measures in each phenotypic group of eye and hair color. It is possible to observe that the mean are very similar between right and left eye. It is also indicate a high consistence in measures.

It was also possible to observe some borderline associations in Table 1. The *MC1R*rs1805009, and *OCA2*rs1800401 showed significant association only with Saturation measure in hair color. The Saturation reflects the ratio of the amount of color compared to medium gray color. How much less gray in the composition of pure color, the more saturated it is. Therefore, it is a measure subject to interference from the natural light of the environment or the flash of the camera. This may explain the association isolated from the two markers mentioned above. And it may also explain the inconsistence of association with the Hue and Saturation in eye and hair color of the three markers more consistent in our analysis (*HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982). As observed in Supplementary Table 5, H and S measures vary from 0 to 1. To reduce RBG mean (from 0 to 255) for H/S could have lost a handful of information in analysis, which can also explain the inconsistent associations. Besides that, the Red and Blond hair is based in only 4 and 27 subjects, respectively, which may to limit our findings.

3.3 Linkage disequilibrium analysis

In the analysis of linkage disequilibrium (LD) for polymorphisms in the same gene and between markers of *OCA2* and *HERC2* genes, considering pairwise analysis made by MLocus software, we observed significant LD between following SNPs: *SLC45A2*rs26722-*SLC45A2*rs16891982, *SLC45A2*rs16891982-*SLC45A2*rs6867641, *TPCN2*rs3750965-*TPCN2*rs3829241, and *TYR*rs1126809-*TYR*rs1042602 (Table 2). The LD for these polymorphisms remained significant after separate analysis of the two major subgroups of the sample, namely: only individuals of Rio Grande do Sul state (sample with predominantly European ancestry), and excluding individuals from Rio Grande do Sul (sample with higher levels of Amerindian and African genetic admixture) (data not shown).

3.4 Graphical analysis of RGB values between eye and hair phenotypes

We produce a 3D graphic in R software to verify if it is possible visually to distinguish the RGB values in individuals that have different phenotype for eye and hair color (Figure 1). Interestingly, a clear quantitative profile of RGB values was made between subjects with Blue/grey and Dark brown/Black eyes (Figure 1a). There was fairly overlap of quantitative values from RGB to other phenotypic classes for eye color (for instance, Green, Honey, and Light brown eyes). This information corroborates the genetic data in the scientific literature, which indicate a more feasibility in to predict Blue or Dark brown/Black eye color. To hair color, we do not observe a clear quantitative differentiation between RGB values (Figure 1b). It is important to note that our data present only 4 subjects with Red/reddish hair color.

4. Discussion

The main implication of our study is the establishment of some genetic markers as eye and hair color predictors to FDP technology in an admixed population. Our analysis presented 3 SNPs (*HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982) significantly associated with eye and hair color in all analysis done. The association between quantitative analysis was quite similar to qualitative

analysis, indicating that phenotypes well categorized are also reliable to present markers associated with physical features. Furthermore, we observe a quantitative profile of RGB values between two eye phenotypic categories. This corroborates literature data indicating a more predictability to forensic practice in the extreme phenotypes considering eye color.

The exclusion criteria to the quantitative analyses of eye and hair color were: individuals using contact lens or suspected to use, observed only in the photograph analyzed; and high level of whitening hair by age or artificially changing pigmentation phenotypes (hair dying). Despite of our rigorous methodology in recruiting volunteers for eye and hair color analysis, it is important to note that there is the possibility that we have included some of those subjects in the quantitative analysis, since only by photo, we were unable to distinguish whether the hair color was considered natural or artificial to some volunteers. We believe that this observation does not invalidate our results, since we believe that there were few cases contemplated in the above observation. Anyway, the most consistent markers in our qualitative and quantitative analyses were *HERC2*rs1129038, *SLC24A5*rs1426654, and *SLC45A2*rs16891982. We note that these three SNPs also presented significantly associated with skin color in our previous data.[29]

All polymorphisms studied here were chosen based on functional data and association studies found in the literature[29], mainly for Europeans and/or Asians populations. It is possible that some differences may exist between different admixed and non-admixed populations, with respect to the association of pigmentation markers analyzed, as noted in previous studies.[9,29] This indicates that some adjustments must be made on the set of SNPs for specific population groups, reducing possible noise in prediction analysis. Walsh *et al.*[23] shown a multiplex assay of 24 polymorphisms denominate HIrisplex for simultaneously to predict eye and hair color. This study is the first prediction system simultaneous of eye and hair color by DNA, opening a new era of forensic genetics. In our data, 6 markers studied by Walsh *et al.*[23] were also analyzed here: *MC1R*rs1805007, *MC1R*rs1805008, *MC1R*rs1805009, *OCA2*rs1800407, *SLC45A2*rs16891982, and *TYR*rs1042602. Of these markers, only one (*SLC45A2*rs16891982) was significantly associated in our sample. This data again emphasizes the fact that genotype-phenotype relationships found in one population cannot be uncritically extrapolated to other populations, as suggested by Cerqueira *et al.*[29] Further analysis must be conduced to reinforce the associations found here, mainly

in admixed groups, and also to check other good markers associated in admixed populations.

Another important genetic marker for eye and hair color verified in our study is *HERC2*rs1129038, which is in high linkage disequilibrium with the HIrisplex's SNP *HERC2*rs12913832 (ref. 36). Ruiz *et al.*[37] have suggested that both SNPs may have a relevant and independent effect in pigmentation pathway, indicating that both can be used in the same prediction set. The third relevant SNP studied here is *SLC24A5*rs1426654. This polymorphism is a novel finding and also seems very promising for FDP in admixed samples. Recently, Liu *et al.*[38] mentioned that *SLC24A5*rs1426654 unlikely contribute to eye and hair human color variation, since the Thr111 allele of this SNP is largely fixed in Europeans, whereas the ancestral Ala111 allele had a high frequency (93 to 100%) in African.[39] Liu *et al.*[38] suggested further investigation of this DNA variant including admixed individuals. Our data showed this SNP highly associated with eye, hair, and also with skin color.[29] As we said, the *SLC24A5*rs1426654 is shown as excellent candidate to be applied to the FDP in admixed sample. In present study, the Ala111 frequency was 15%. Besides of the small number of subjects with Red/reddish and Blond hair color, compared with other phenotypes, as well as the small number of subjects with Blue/grey and Green eye color, the significance of this three significant markers in our analysis was quite high.

To the best of our knowledge, Neitzke-Montinelli *et al.*[40] was the first Brazilian study analyzing normal variation polymorphisms associated with human pigmentation. These authors analyzed two SNPs in the *MC1R* gene (rs3212345 and rs3212346), and found an association with skin and hair pigmentation, as well as tanning ability. An interesting fact in our work is the lack of association of classical markers with eyes (i.e. *OCA2* gene markers) and hair (i.e. *MC1R* gene markers) color in the sample admixed studied here. The three common *MC1R* SNPs studied (rs1805007, rs1805008, and rs1805009) are classics in the pigmentation pathway, at least considering European populations, being commonly associated with red hair pigmentation[13, 19, 24, 41, 42], although it is believed that they also account for substantial variation in skin color.[19, 24, 43, 44] The combination of the alleles in these SNPs results in a range of hair color shades.[45, 46] Our work not showed such association. The small quantity of red hair subjects in our sample could explain this issue. With relation *OCA2* gene SNPs (*OCA2*rs1800401, *OCA2*rs1800407 and *OCA2*rs1800414), our data also indicated none relationship between these markers and eye color in Brazilians. An alternative

explanation is that the *MC1R* and *OCA2* polymorphisms studied are almost fixed in our admixed samples (please, consult Supplementary Table 2). Further studies must be done to verify in more details the *MC1R* and *OCA2* effects in admixed samples.

According Liu *et al.*[38], future studies should concentrate on high accuracy and high level of detail in color phenotype description. Here, we consider this recommendation. Despite the efforts of scientific community in quantify phenotypes to the better phenotypic prediction, we observed that the use of qualitative data shown the same results, compared with quantitative data. It is known that categorized information inevitably oversimplifies continuous traits.[38] However, the quantitative data introduce a continuous spectrum that is difficult to define what the cutoff points needed to define the phenotypes in question. Qualitative data, when well defined, could avoid these concerns. We believe that it should be better discussed in the scientific community.

Finally, one advantage of studying the LD between genetic markers for forensic application is to have available some SNPs at hand for analysis with incomplete profiles in usual sets, in which it was not possible to genotype some markers[37], as in mass disasters. We present some LD between SNPs that may be useful for this purpose. Unfortunately, some these LD markers are not associated with eye or hair phenotypes in present study.

Conflict of interest

The authors declare no conflict of interest

Acknowledgments

We are grateful to the sponsors CAPES, CNPq, FAPERGS, The Leverhulme Trust, and The Biotechnology and Biological Sciences Research Council. Finally, we would like to thank the Executive Commitee members and especially to the volunteers of the CANDELA Project, for making this study possible.

References

[1] Davenport GC, Davenport CB: Heredity of eye-color in man. *Science* 1907; **26**: 590–592.

[2] Hurst CC: On the inheritance of eye-colour in man. *Proc R Soc Med Lond B Biol Sci* 1908; **80**: 85–96.

[3] Holmes SJ, Loomis HM: The heredity of eye color and hair color in man. *Biol Bull* 1909; **18**: 50–65.

[4] Wright S: Color inheritance in mammals. *J. Hered* 1918; **9**: 227–240.

[5] Yamaguchi Y, Hearing VJ: Physiological factors that regulate skin pigmentation. *Biofactors* 2009; **35**: 193-199.

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

[7] Graf J, Hodgson R, Van Daal A: Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. *Hum. Mutat* 2005; **25**: 278–284.

[8] Graf J, Voisey J, Hughes I, van Daal, A: Promoter polymorphisms in the MATP (SLC45A2) gene are associated with normal human skin color variation. *Hum Mut* 2007; **28**: 710-717.

[9] Norton HL, Kittles RA, Parra E, *et al*: Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol Biol Evol* 2007; **24**: 710–722.

[10] Stokowski RP, Pant PV, Dadd T, *et al*: A genomewide association study of skin pigmentation in a South Asian population. *Am J Hum Genet* 2007; **81**: 1119-1132.

[11] Edwards M, Bigham A, Tan J, *et al*: Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: Further evidence of convergent evolution of skin pigmentation. *PLoS Genet* 2010; **6**: e1000867.

[12] Price AL, Patterson N, Yu F, *et al*: A genomewide admixture map for Latino populations. *Am J Hum Genet* 2007; **80**: 1024-1036.

[13] Sulem P, Gudbjartsson DF, Stacey SN, *et al*: Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 2007; **39**: 1443-1452.

[14] Marrero AR, Bravi C, Stuart S, *et al*: Pre- and post-Columbian gene and cultural continuity: the case of the Gaucho from southern Brazil. *Hum Hered* 2007; **64**: 160-171.

[15] Hünemeier T, Carvalho C, Marrero AR, Salzano FM, Junho Pena SD, Bortolini MC: Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. *Am J Phys Anthropol* 2007; **133**: 854-867.

[16] Guerreiro-Junior V, Bisso-Machado R, Marrero A, Hünemeier T, Salzano FM, Bortolini MC: Genetic signatures of parental contribution in black and white populations in Brazil. *Genet Mol Biol* 2009; **32**: 1-11.

[17] Draus-Barini J, Walsh S, Pośpiech E, *et al*: Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investig Genet* 2013; **4**: 3.

[18] Valverde P, Healy E, Jackson I, Rees J, Thody AJ: Variants of the melanocyte stimulating hormone receptor gene are associated with red hair and fair skin. *Nat Genet* 1995; **11**: 328–330.

[19] Duffy DL, Box NF, Chen W, *et al*: Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum Mol Genet* 2004; **13**: 447-461.

[20] Jablonski NG, Chaplin G: The evolution of human skin coloration. *J Hum Evol* 2000; **39**: 57–106.

[21] Parra EJ: Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am J Phys Anthropol* 2007; **45**: 85-105.

[22] Kayser M, Knijff P: Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet* 2011; **12**: 179–192.

[23] Walsh S, Liu F, Wollstein A, *et al*: The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet* 2013; **7**: 98-115.

[24] Sulem P, Gudbjartsson DF, Stacey SN, *et al*: Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet* 2008; **40**: 835–837.

[25] Liu F, van Duijn K, Vingerling JR, *et al*: Eye color and the prediction of complex phenotypes from genotypes. *Curr Biol* 2009; **19**: R192-R193.

[26] Valenzuela RK, Henderson MS, Walsh MH, *et al*: Predicting phenotype from genotype: normal pigmentation. *J Forensic Sci* 2010; **55**: 315-322.

[27] Branicki W, Liu F, van Duijn K, *et al*: Model-based prediction of human hair color using DNA variants. *Hum Genet* 2011; **129**: 443–454.

[28] Kayser M, Schneider PM: DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. *Forensic Sci Int Genet* 2009; **3**: 154–161.

[29] Cerqueira CCS, Hünemeier T, Gomez-Valdes JA, for CANDELA Consortium, *et al*: Forensic implications of the genetics evaluation of skin color phenotype in two distinct admixed Brazilian populations. *PLoS One* 2013; SUBMETIDO.

[30] Bastos JL, Dumith SC, Santos RV, *et al*: Does the way I see you affect the way I see myself? Associations between interviewers' and interviewees' "color/race" in southern Brazil. *Cad Saúde Pública* 2009; **25**: 2111-2124.

[31] Duffy DL, Montgomery GW, Chen W, *et al*: A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am J Hum Genet* 2007; **80**: 241–252.

[32] Eiberg H, Troelsen J, Nielsen M, *et al*: Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet* 2008; **123**:177–187.

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

[34] Liu F, Wollstein A, Hysi PG, *et al*: Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet* 2010; **6**: e1000934.

[35] Long, JC: Multiple Locus Haplotype Analysis, version 3.0. Software and documentation distributed by the author Department of Human Genetics, University of Michigan Medical School, 1999; Ann Arbor, MI 4819-0618.

[36] Amos CI, Wang L, Lee JE, *et al*: Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet* 2011; **20**: 5012–5023.

[37] Ruiz Y, Phillips C, Gomez-Tato A, *et al*: Further development of forensic eye color predictive tests. *Forensic Sci Int Genet* 2013; **7**: 28-40.

[38] Liu F, Wen B, Kayser M: Colorful DNA polymorphisms in humans. *Semin Cell Dev Biol* 2013; **24**: 562-575.

[39] Lamason RL, Mohideen MA, Mest JR, *et al*: SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 2005; **310**:1782-1786.

[40] Neitzke-Montinelli V, Urmenyi TP, Rondinelli E, Cabello PH, Silva R, Moura-Neto RS: Polymorphisms upstream of the melanocortin-1 receptor coding region are associated with human pigmentation variation in a Brazilian population. *Am J Hum Biol* 2012; **24**: 853-855.

[41] Smith R, Healy E, Siddiqui S, *et al*: Melanocortin 1 receptor variants in an Irish population. *J Invest Dermatol* 1998; **111**: 119-122.

[42] Latreille J, Ezzedine K, Elfakir A, *et al*: MC1R gene polymorphism affects skin color and phenotypic features related to sun sensitivity in a population of French adult women. *Photochem Photobiol* 2009; **85**: 1451-1458.

[43] Rees JL: The genetics of sun sensitivity in humans. *Am J Hum Genet* 2004; **75**: 739-751.

[44] Makova K, Norton H: Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. *Peptides* 2005; **26**: 1901–1908.

[45] Rees J, Birch-Machin M, Flanagan N, Healy E, Phillips S, Tood C: Genetic studies of the human melanocortin-1 receptor. *Ann NY Acad Sci* 1999; **885**: 143–152.

[46] Flanagan N, Healy E, Ray A, *et al*: Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. *Hum Mol Genet* 2000; **9**: 2531-2537.

Table 1. Summary of p values of all analyses with 18 markers and eye/ hair color in our sample.

SNPs	**Eye color**					
	Qualitative analyses	**Quantitative analyses**				
		Red	**Green**	**Blue**	**Hue**	**Saturation**
*ADAM17*rs1524668	0.814	0.565	0.198	0.108	0.067*	0.259
*AFG3L1*rs4785763	0.153	0.455	0.133	0.125*	0.335*	0.126*
*ASIP*rs6058017	0.514	0.471	0.112*	0.171*	0.001*	0.331
*HERC2*rs1129038	**<0.001**	**<0.001***	**<0.001***	**<0.001***	**<0.001***	**<0.001**
*MC1R*rs1805007	0.225	0.858	0.865	0.898	0.828	0.268
*MC1R*rs1805008	0.459	0.188	0.593	0.834	0.801	0.569
*MC1R*rs1805009	0.457	0.478	0.361	0.367	0.988	0.390
*OCA2*rs1800401	0.938	0.142	0.151	0.275	0.195	0.875
*OCA2*rs1800407	0.536	0.062	0.282	0.649	0.293*	0.075
*OCA2*rs1800414	0.876	0.171	0.183	0.364	0.321	0.512*
*SLC24A5*rs1426654	**<0.001**	**<0.001***	**<0.001***	**<0.001***	**<0.001***	0.129
*SLC45A2*rs16891982	**<0.001**	**<0.001**	**<0.001***	**<0.001***	**<0.001***	0.148*
*SLC45A2*rs26722	0.322	0.396*	0.356*	0.448*	0.163*	0.439
*SLC45A2*rs6867641	0.696	0.142	0.192	0.195	0.573*	0.981
*TPCN2*rs3750965	0.499	0.948	0.853	0.784	0.904	0.553
*TPCN2*rs3829241	0.276	0.130	0.093*	0.560*	0.004*	0.971
*TYR*rs1042602	0.200	0.004	0.052	0.434	0.487	0.134
*TYR*rs1126809	0.002	0.075	0.005	0.008*	0.001*	0.204
SNPs	**Hair color**					
	Qualitative analyses	**Quantitative analyses**				
		Red	**Green**	**Blue**	**Hue**	**Saturation**
*ADAM17*rs1524668	0.437	0.365	0.285	0.157	0.598*	0.929
*AFG3L1*rs4785763	0.156	0.048	0.050	0.135	0.835	0.524*
*ASIP*rs6058017	0.577	0.049	0.078	0.220	0.277*	0.018

*HERC2*rs1129038	**<0.001**	**<0.001***	**<0.001***	**<0.001***	0.006*	**<0.001***
*MC1R*rs1805007	0.018	0.022*	0.021*	0.016*	0.847	0.024*
*MC1R*rs1805008	0.133	0.819	0.561	0.279	0.700	0.289
*MC1R*rs1805009	0.037	0.006*	0.016*	0.127	0.229*	**<0.001**
*OCA2*rs1800401	0.346	0.002*	0.023	0.076	0.431*	**<0.001***
*OCA2*rs1800407	0.240	0.199	0.201	0.230	0.712	0.116
*OCA2*rs1800414	0.471	0.803	0.894	0.437	0.474	0.228
*SLC24A5*rs1426654	**<0.001**	**<0.001***	**<0.001***	**<0.001***	0.543*	**<0.001***
*SLC45A2*rs16891982	**<0.001**	**<0.001***	**<0.001***	**<0.001***	0.098*	**<0.001***
*SLC45A2*rs26722	0.009	0.032	0.004*	0.005*	0.960	0.058
*SLC45A2*rs6867641	0.882	0.984	0.971	0.467	0.111	0.693
*TPCN2*rs3750965	0.177	0.607	0.557	0.576	0.869	0.491
*TPCN2*rs3829241	0.042	0.218	0.366	0.900	0.616*	0.003
*TYR*rs1042602	0.006	0.026	0.034	0.096	0.760*	0.006
*TYR*rs1126809	0.002	0.049*	0.100*	0.106*	0.312*	0.026

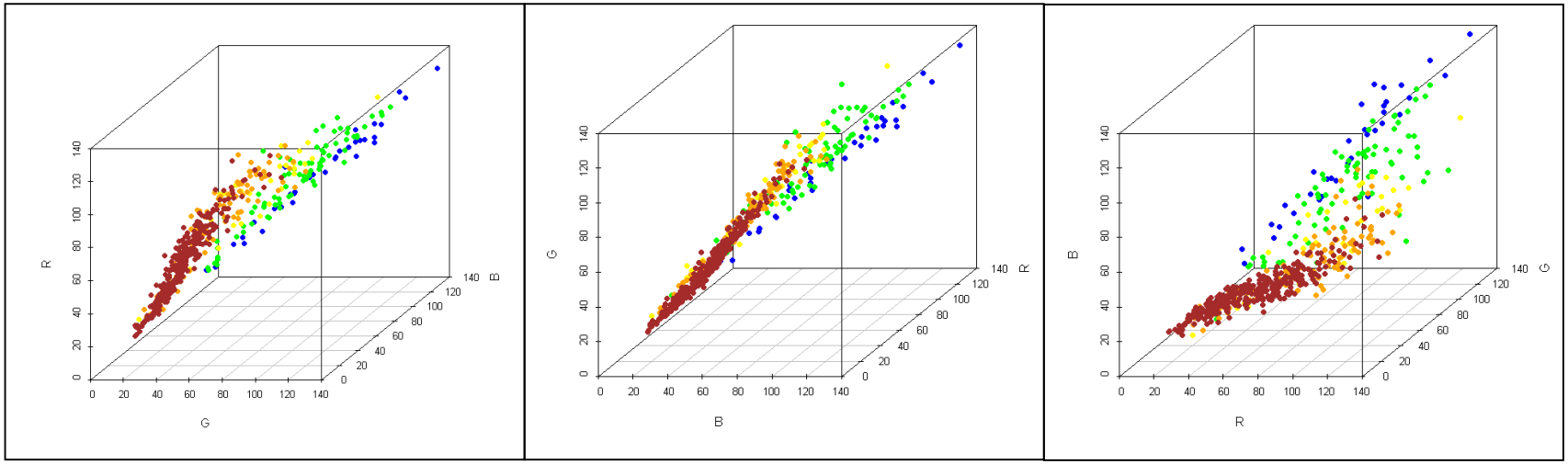
* Non-parametric Kruskal-Wallis test. All the others quantitative analyses were done with ANOVA.

Table 2. Linkage disequilibrium analysis between some polymorphisms studied in this work.

MC1R	**p**	**D'**	**R^2**
rs1805008-rs1805007	0.359	1.000	0.001
rs1805008-rs1805009	0.471	1.000	0.000
rs1805007-rs1805009	0.249	0.044	0.001
OCA2	**p**	**D'**	**R^2**
rs1800407-rs1800401	0.016	1.000	0.005
rs1800407-rs1800414	0.605	1.000	0.000
rs1800401-rs1800414	0.572	1.000	0.000
OCA2-HERC2	**p**	**D'**	**R^2**
rs1800401-rs1129038	0.011	0.376	0.006
rs1800407-rs1129038	0.079	0.283	0.003
rs1800414-rs1129038	0.146	1.000	0.002
SLC45A2	**p**	**D'**	**R^2**
rs26722-rs16891982	**<0.001**	**0.981**	**0.217**
rs26722-rs6867641	0.705	0.026	0.000
rs16891982-rs6867641	0.002	0.103	0.009
TPCN2	**p**	**D'**	**R^2**
rs3750965-rs3829241	**<0.001**	**0.985**	**0.193**
TYR	**p**	**D'**	**R^2**
rs1126809-rs1042602	**<0.001**	**0.740**	**0.067**

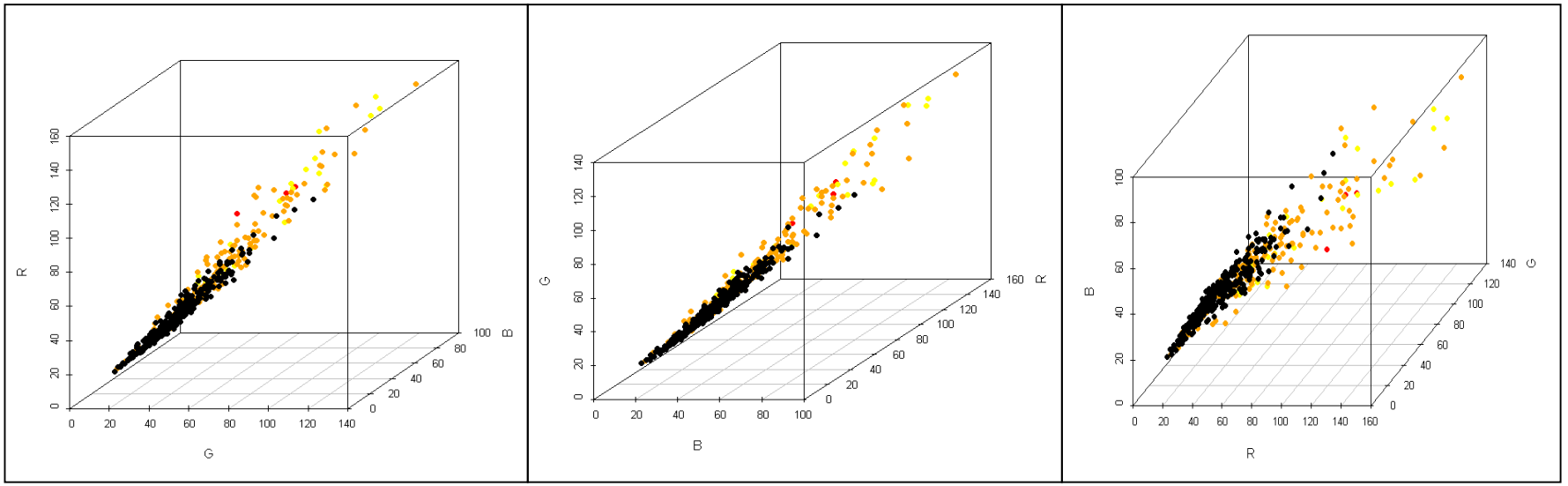
Figure 01. Tridimensional graphic of RGB values, separated by qualitative phenotypic classes of eye and hair color.

a) Graphic RGB in R software to eye color groups.



Legend: Blue= blue/grey eyes; Green= green eyes; Yellow= honey eyes; Orange=light brown eyes; Red= dark brown/black eyes.

b) Graphic RGB in R software to hair color groups.



Legend: Red= red/reddish hair; Yellow= blond hair; Orange= dark blond/light brown hair; Black= dark brown/black hair.

Supplementary Table 1. General characteristics of the studied sample.

	Mean	(Minimum-Maximum)
Age	27.74	(18-85)
	N	**(%)**
Gender		
Male	195	(34.6%)
Female	368	(65.4%)
Eye color		
Blue/Grey	35	(6.2%)
Green	77	(13.7%)
Honey	35	(6.2%)
Light brown	99	(17.6%)
Dark brown/Black	317	(56.3%)
Hair color		
Red/reddish	4	(0.7%)
Blond	34	(6%)
Dark blond/Light Brown	177	(31.4%)
Dark brown/Black	348	(61.8%)

Supplementary Table 2. Genotype and allele frequencies for 18 SNPs in our sample.

	rs1524668	**rs4785763**	**rs6058017**	**rs1129038**	**rs1805007**	**rs1805008**
	1=G/2=T	1=A/2=C	1=G/2=A	1=G/2=A	1=C/2=T	1=C/2=T
Genotypes/Alleles						
11	0.405	0.050	0.034	0.451	0.962	0.925
12	0.469	0.353	0.268	0.411	0.038	0.075
22	0.126	0.597	0.698	0.138	0.000	0.000
1	0.64	0.23	0.17	0.66	0.98	0.96
2	0.36	0.77	0.83	0.34	0.02	0.04
	rs1805009	**rs1800401**	**rs1800407**	**rs1800414**	**rs1426654**	**rs16891982**
	1=G/2=C	1=C/2=T	1=G/2=A	1=A/2=G	1=G/2=A	1=C/2=G
Genotypes/Alleles						
11	0.976	0.860	0.876	0.993	0.045	0.137
12	0.024	0.134	0.124	0.007	0.205	0.345
22	0.000	0.006	0.000	0.000	0.750	0.518
1	0.99	0.93	0.94	0.99	0.15	0.31
2	0.01	0.07	0.06	0.01	0.85	0.69
	rs26722	**rs6867641**	**rs3750965**	**rs3829241**	**rs1042602**	**rs1126809**
	1=C/2=T	1=C/2=T	1=A/2=G	1=G/2=A	1=C/2=A	1=G/2=A
Genotypes/Alleles						
11	0.823	0.457	0.468	0.511	0.411	0.696
12	0.173	0.440	0.419	0.389	0.434	0.265
22	0.004	0.103	0.113	0.100	0.155	0.039
1	0.91	0.68	0.68	0.71	0.63	0.83
2	0.09	0.32	0.32	0.29	0.37	0.17

Supplementary Table 3. Detailed association analysis between qualitative eye color and genotypes.

Gene	SNPs	Genotype (N)	Eye color					χ^{2*}	p
			Blue/Grey	Green	Honey	Light brown	Darker brown/Black		
ADAM17	rs1524668	GG (223)	17 (48.6%)	36 (48%)	15 (44.1%)	37 (38.5%)	118 (38.1%)		
		GT (258)	15 (42.9%)	29 (38.7%)	16 (47.1%)	47 (49%)	151 (48.7%)		
		TT (69)	3 (8.6%)	10 (13.3%)	3 (8.8%)	12 (12.5%)	41 (13.2%)		
			35 (100%)	75 (100%)	34 (100%)	96 (100%)	310 (100%)	4.505	0.814
AFG3L1	rs4785763	AA (27)	2 (5.9%)	6 (8.6%)	1 (2.9%)	3 (3.1%)	15 (5%)		
		AC (190)	16 (47.1%)	32 (45.7%)	12 (35.3%)	30 (30.9%)	100 (33%)		
		CC (321)	16 (47.1%)	32 (45.7%)	21 (61.8%)	64 (66%)	188 (62%)		
			34 (100%)	70 (100%)	34 (100%)	97 (100%)	303 (100%)	11.313	0.153
ASIP	rs6058017	GG (18)	0 (0%)	0 (0%)	1 (3%)	3 (3.2%)	14 (4.6%)		
		GA (144)	7 (20.6%)	17 (23.3%)	8 (24.2%)	29 (30.5%)	83 (27.4%)		
		AA (376)	27 (79.4%)	56 (76.7%)	24 (72.7%)	63 (66.3%)	206 (68%)		
			34 (100%)	73 (100%)	33 (100%)	95 (100%)	303 (100%)	6.827	0.514
HERC2	rs1129038	GG (173)	0 (0%)	1 (2.1%)	2 (9.5%)	21 (30.4%)	149 (66.2%)		
		GA (158)	4 (19%)	19 (39.6%)	14 (66.7%)	46 (66.7%)	75 (33.3%)		
		AA (53)	17 (81%)	28 (58.3%)	5 (23.8%)	2 (2.9%)	1 (0.4%)		
			21 (100%)	48 (100%)	21 (100%)	69 (100%)	225 (100%)	**228.419**	**<0.001**
MC1R	rs1805007	CC (535)	34 (97.1%)	71 (93.4%)	32 (91.4%)	97 (98%)	301 (96.8%)		
		CT (21)	1 (2.9%)	5 (6.6%)	3 (8.6%)	2 (2%)	10 (3.2%)		
		TT (0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
			35 (100%)	76 (100%)	35 (100%)	99 (100%)	311 (100%)	4.983	0.225
MC1R	rs1805008	CC (172)	13 (86.7%)	28 (87.5%)	12 (92.3%)	32 (97%)	87 (93.5%)		
		CT (14)	2 (13.3%)	4 (12.5%)	1 (7.7%)	1 (3%)	6 (6.5%)		
		TT (0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
			15 (100%)	32 (100%)	13 (100%)	33 (100%)	93 (100%)	3.387	0.459

MC1R	rs1805009	GG	(538)	32 (97%)	73 (96.1%)	32 (97%)	95 (96.9%)	306 (98.4%)		
		GC	(13)	1 (3%)	3 (3.9%)	1 (3%)	3 (3.1%)	5 (1.6%)		
		CC	(0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
				33 (100%)	76 (100%)	33 (100%)	98 (100%)	311 (100%)	3.180	0.457
OCA2	rs1800401	CC	(473)	32 (91.4%)	65 (87.8%)	31 (91.2%)	84 (85.7%)	261 (84.5%)		
		CT	(74)	3 (8.6%)	9 (12.2%)	3 (8.8%)	14 (14.3%)	45 (14.6%)		
		TT	(3)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (1%)		
				35 (100%)	74 (100%)	34 (100%)	98 (100%)	309 (100%)	3.074	0.938
OCA2	rs1800407	GG	(478)	32 (91.4%)	60 (83.3%)	32 (94.1%)	84 (85.7%)	270 (87.9%)		
		GA	(68)	3 (8.6%)	12 (16.7%)	2 (5.9%)	14 (14.3%)	37 (12.1%)		
		AA	(0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
				35 (100%)	72 (100%)	34 (100%)	98 (100%)	307 (100%)	3.068	0.536
OCA2	rs1800414	AA	(549)	35 (100%)	76 (100%)	34 (100%)	99 (100%)	305 (98.7%)		
		AG	(4)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (1.3%)		
		GG	(0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
				35 (100%)	76 (100%)	34 (100%)	99 (100%)	309 (100%)	1.642	0.876
SLC24A5	rs1426654	GG	(25)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	25 (8%)		
		GA	(114)	2 (5.7%)	5 (6.6%)	4 (11.8%)	11 (11.1%)	92 (29.6%)		
		AA	(416)	33 (94.3%)	71 (93.4%)	30 (88.2%)	88 (88.9%)	194 (62.4%)		
				35 (100%)	76 (100%)	34 (100%)	99 (100%)	311 (100%)	**63.720**	**<0.001**
SLC45A2	rs16891982	CC	(74)	0 (0%)	3 (4.1%)	1 (3%)	9 (9.2%)	61 (20.3%)		
		CG	(187)	3 (8.6%)	21 (28.4%)	11 (33.3%)	25 (25.5%)	127 (42.2%)		
		GG	(280)	32 (91.4%)	50 (67.6%)	21 (63.6%)	64 (65.3%)	113 (37.5%)		
				35 (100%)	74 (100%)	33 (100%)	98 (100%)	301 (100%)	**70.856**	**<0.001**
SLC45A2	rs26722	CC	(453)	32 (91.4%)	62 (84.9%)	26 (76.5%)	86 (86.9%)	247 (79.9%)		
		CT	(95)	3 (8.6%)	11 (15.1%)	8 (23.5%)	12 (12.1%)	61 (19.7%)		
		TT	(2)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	1 (0.3%)		
				35 (100%)	73 (100%)	34 (100%)	99 (100%)	309 (100%)	8.905	0.322

SLC45A2	rs6867641	CC	(84)	9 (64.3%)	13 (39.4%)	5 (41.7%)	14 (42.4%)	43 (46.7%)		
		CT	(81)	3 (21.4%)	16 (48.5%)	7 (58.3%)	15 (45.5%)	40 (43.5%)		
		TT	(19)	2 (14.3%)	4 (12.1%)	0 (0%)	4 (12.1%)	9 (9.8%)		
				14 (100%)	33 (100%)	12 (100%)	33 (100%)	92 (100%)	5.543	0.696
TPCN2	rs3750965	AA	(256)	16 (45.7%)	33 (44.6%)	14 (41.2%)	39 (40.2%)	154 (50.2%)		
		AG	(229)	12 (34.3%)	34 (45.9%)	17 (50%)	46 (47.4%)	120 (39.1%)		
		GG	(62)	7 (20%)	7 (9.5%)	3 (8.8%)	12 (12.4%)	33 (10.7%)		
				35 (100%)	74 (100%)	34 (100%)	97 (100%)	307 (100%)	7.268	0.499
TPCN2	rs3829241	GG	(276)	18 (51.4%)	30 (41.1%)	14 (43.8%)	49 (50%)	165 (54.6%)		
		GA	(210)	12 (34.3%)	37 (50.7%)	13 (40.6%)	36 (36.7%)	112 (37.1%)		
		AA	(54)	5 (14.3%)	6 (8.2%)	5 (15.6%)	13 (13.3%)	25 (8.3%)		
				35 (100%)	73 (100%)	32 (100%)	98 (100%)	302 (100%)	9.810	0.276
TYR	rs1042602	CC	(221)	12 (35.3%)	26 (36.1%)	13 (40.6%)	31 (32%)	139 (46%)		
		CA	(233)	17 (50%)	36 (50%)	12 (37.5%)	45 (46.4%)	123 (40.7%)		
		AA	(83)	5 (14.7%)	10 (13.9%)	7 (21.9%)	21 (21.6%)	40 (13.2%)		
				34 (100%)	72 (100%)	32 (100%)	97 (100%)	302 (100%)	10.833	0.200
TYR	rs1126809	GG	(380)	17 (51.5%)	46 (62.2%)	19 (57.6%)	60 (63.8%)	238 (76.3%)		
		GA	(145)	12 (36.4%)	25 (33.8%)	11 (33.3%)	30 (31.9%)	67 (21.5%)		
		AA	(21)	4 (12.1%)	3 (4.1%)	3 (9.1%)	4 (4.3%)	7 (2.2%)		
				33 (100%)	74 (100%)	33 (100%)	94 (100%)	312 (100%)	22.552	0.002

* Fisher's Exact test.

Supplementary Table 4. Detailed association analysis between qualitative hair color and genotypes.

Gene	SNPs	Genotype (N)	Hair color				χ^{2*}	p
			Red/Reddish	Blond	Dark blond/ Light brown	Dark brown/ Black		
ADAM17	rs1524668	GG (223)	2 (50%)	18 (52.9%)	62 (36%)	141 (41.5%)		
		GT (258)	2 (50%)	13 (38.2%)	91 (52.9%)	152 (44.7%)		
		TT (69)	0 (0%)	3 (8.8%)	19 (11%)	47 (13.8%)		
			4 (100%)	34 (100%)	172 (100%)	340 (100%)	5.599	0.437
AFG3L1	rs4785763	AA (27)	0 (0%)	4 (11.8%)	7 (4.1%)	16 (4.8%)		
		AC (190)	2 (50%)	15 (44.1%)	67 (39.4%)	106 (32.1%)		
		CC (321)	2 (50%)	15 (44.1%)	96 (56.5%)	208 (63%)		
			4 (100%)	34 (100%)	170 (100%)	330 (100%)	8.812	0.156
ASIP	rs6058017	GG (18)	0 (0%)	0 (0%)	5 (2.9%)	13 (3.9%)		
		GA (144)	0 (0%)	6 (18.2%)	44 (25.7%)	94 (28.5%)		
		AA (376)	4 (100%)	27 (81.8%)	122 (71.3%)	223 (67.6%)		
			4 (100%)	33 (100%)	171 (100%)	330 (100%)	4.461	0.577
HERC2	rs1129038	GG (173)	0 (0%)	0 (0%)	29 (25.7%)	144 (58.1%)		
		GA (158)	2 (100%)	5 (23.8%)	60 (53.1%)	91 (36.7%)		
		AA (53)	0 (0%)	16 (76.2%)	24 (21.2%)	13 (5.2%)		
			2 (100%)	21 (100%)	113 (100%)	248 (100%)	**93.498**	**<0.001**
MC1R	rs1805007	CC (535)	3 (75%)	31 (91.2%)	166 (94.9%)	335 (97.7%)		
		CT (21)	1 (25%)	3 (8.8%)	9 (5.1%)	8 (2.3%)		
		TT (0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
			4 (100%)	34 (100%)	175 (100%)	343 (100%)	9.643	0.018
MC1R	rs1805008	CC (172)	2 (100%)	11 (78.6%)	60 (90.9%)	99 (95.2%)		
		CT (14)	0 (0%)	3 (21.4%)	6 (9.1%)	5 (4.8%)		
		TT (0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
			2 (100%)	14 (100%)	66 (100%)	104 (100%)	5.358	0.133

MC1R	rs1805009	GG (538)	3 (75%)	31 (93.9%)	168 (97.7%)	336 (98.2%)		
		GC (13)	1 (25%)	2 (6.1%)	4 (2.3%)	6 (1.8%)		
		CC (0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
			4 (100%)	33 (100%)	172 (100%)	342 (100%)	8.125	0.037
OCA2	rs1800401	CC (473)	4 (100%)	33 (97.1%)	150 (86.2%)	286 (84.6%)		
		CT (74)	0 (0%)	1 (2.9%)	24 (13.8%)	49 (14.5%)		
		TT (3)	0 (0%)	0 (0%)	0 (0%)	3 (0.9%)		
			4 (100%)	34 (100%)	174 (100%)	338 (100%)	7.338	0.346
OCA2	rs1800407	GG (478)	4 (100%)	30 (90.9%)	144 (83.2%)	300 (89.3%)		
		GA (68)	0 (0%)	3 (9.1%)	29 (16.8%)	36 (10.7%)		
		AA (0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
			4 (100%)	33 (100%)	173 (100%)	336 (100%)	4.002	0.240
OCA2	rs1800414	AA (549)	4 (100%)	34 (100%)	177 (100%)	334 (99%)		
		AG (4)	0 (0%)	0 (0%)	0 (0%)	4 (1%)		
		GG (0)	0 (0%)	0 (0%)	0 (0%)	0 (0%)		
			4 (100%)	34 (100%)	177 (100%)	338 (100%)	3.651	0.471
SLC24A5	rs1426654	GG (25)	0 (0%)	0 (0%)	1 (0.6%)	24 (7%)		
		GA (114)	1 (33.3%)	1 (2.9%)	17 (9.7%)	95 (27.8%)		
		AA (416)	2 (66.7%)	33 (97.1%)	158 (89.8%)	223 (65.2%)		
			3 (100%)	34 (100%)	176 (100%)	342 (100%)	**52.423**	**<0.001**
SLC45A2	rs16891982	CC (74)	0 (0%)	0 (0%)	7 (4.1%)	67 (20.2%)		
		CG (187)	2 (50%)	4 (11.8%)	32 (18.7%)	149 (44.9%)		
		GG (280)	2 (50%)	30 (88.2%)	132 (77.2%)	116 (34.9%)		
			4 (100%)	34 (100%)	171 (100%)	332 (100%)	**107.088**	**<0.001**
SLC45A2	rs26722	CC (453)	3 (75%)	32 (94.1%)	154 (88.5%)	264 (78.1%)		
		CT (95)	1 (25%)	2 (5.9%)	19 (10.9%)	73 (21.6%)		
		TT (2)	0 (0%)	0 (0%)	1 (0.6%)	1 (0.3%)		
			4 (100%)	34 (100%)	174 (100%)	338 (100%)	16.762	0.009

SLC45A2	rs6867641	CC (84)	1 (50%)	6 (46.2%)	30 (45.5%)	47 (45.6%)		
		CT (81)	1 (50%)	7 (53.8%)	27 (40.9%)	46 (44.7%)		
		TT (19)	0 (0%)	0 (0%)	9 (13.6%)	10 (9.7%)		
			2 (100%)	13 (100%)	66 (100%)	103 (100%)	2.730	0.882
TPCN2	rs3750965	AA (256)	1 (33.3%)	16 (47.1%)	71 (42%)	168 (49.3%)		
		AG (229)	1 (33.3%)	17 (50%)	80 (47.3%)	131 (38.4%)		
		GG (62)	1 (33.3%)	1 (2.9%)	18 (10.7%)	42 (12.3%)		
			3 (100%)	34 (100%)	169 (100%)	341 (100%)	8.311	0.177
TPCN2	rs3829241	GG (276)	4 (100%)	12 (35.3%)	80 (47.1%)	180 (54.2%)		
		GA (210)	0 (0%)	20 (58.8%)	67 (39.4%)	123 (37%)		
		AA (54)	0 (0%)	2 (5.9%)	23 (13.5%)	29 (8.7%)		
			4 (100%)	34 (100%)	170 (100%)	332 (100%)	12.060	0.042
TYR	rs1042602	CC (221)	1 (25%)	10 (29.4%)	52 (31%)	158 (47.7%)		
		CA (233)	2 (50%)	19 (55.9%)	83 (49.4%)	129 (39%)		
		AA (83)	1 (25%)	5 (14.7%)	33 (19.6%)	44 (13.3%)		
			4 (100%)	34 (100%)	168 (100%)	331 (100%)	16.977	0.006
TYR	rs1126809	GG (380)	3 (75%)	15 (45.5%)	112 (65.9%)	250 (73.7%)		
		GA (145)	0 (0%)	15 (45.5%)	49 (28.8%)	81 (23.9%)		
		AA (21)	1 (25%)	3 (9.1%)	9 (5.3%)	8 (2.4%)		
			4 (100%)	33 (100%)	170 (100%)	339 (100%)	19.323	0.002

* Fisher's Exact test.

Supplementary Table 5. Mean Red, Green, Blue, Hue, and Saturation values in each phenotypic group of eye and hair color.

	Eye color (Right eye/ Left eye)				
	Red Mean	**Green Mean**	**Blue Mean**	**Hue**	**Saturation**
Blue (n=34)	84.19/ 84.98	87.42/ 88.63	90.53/ 91.97	0.502/ 0.486	0.099/ 0.108
Green (n=77)	88.21/ 88.26	83.47/ 83.03	74.61/ 74.34	0.212/ 0.189	0.176/ 0.177
Honey (n=35)	83.80/ 83.26	65.66/ 66.39	49.08/ 50.16	0.081/ 0.085	0.416/ 0.398
Light brown (n=98)	77.72/ 77.50	55.07/ 54.67	40.79/ 40.20	0.063/ 0.063	0.458/ 0.459
Dark brown (n=316)	56.45/ 58.83	38.32/ 39.86	31.85/ 32.74	0.034/ 0.063	0.399/ 0.408
	Hair color				
	Red Mean	**Green Mean**	**Blue Mean**	**Hue**	**Saturation**
Red (n=4)	83.61	62.46	46.68	0.041	0.384
Blond (n=27)	84.41	66.98	51.17	0.076	0.361
Light brown (n=159)	60.24	49.92	42.10	0.066	0.253
Dark brown (n=326)	39.34	35.26	33.83	0.138	0.138

4. DISCUSSÃO

Após a publicação dos rascunhos dos genomas do Neandertal (Green *et al*., 2010) e de um espécime encontrado na caverna de Denisova, Rússia (Reich *et al*., 2010), foram desvendandos pela primeira vez detalhes genéticos e moleculares de hominíneos não-sapiens. Porém, antes destas publicações, alguns pesquisadores tentaram inferir a aparência destes hominíneos extintos. Um dos estudos dignos de menção é o de Lalueza-Fox *et al*. (2007). Os autores sequenciaram um fragmento do gene *MC1R* de dois espécimes de Neandertais e encontraram uma variante possivelmente funcional que estaria associada à perda de função do *MC1R* conferindo, possivelmente, uma pele clara e cabelos ruivos/avermelhados nos indivíduos que a possuem, pois se sabe que cabelos vermelhos é o fenótipo nulo do *MC1R* (Beaumont *et al*., 2008). Posteriormente, Bouakaze *et al*. (2009), numa análise igualmente interessante, genotiparam em um único ensaio multiplex 10 SNPs em 6 genes candidatos para pigmentação com a finalidade de produzir informações sobre traços de pigmentação de 25 restos humanos arqueológicos provenientes da Sibéria, datados da Idade do Bronze e do Ferro (entre 5.000 e 1.000 anos antes do presente). Mais recentemente Draus-Barini *et al*. (2013) tentaram predizer traços de pigmentação de 26 fragmentos ósseos humanos antigos, utilizando o sistema Hirisplex, desenvolvido por Walsh *et al.* (2013) para estimar cor de olhos e cabelos a partir da análise de DNA.

Numa perspectiva similar, a publicação do nosso grupo (Cerqueira *et al*., 2012 – Capítulo I dos resultados desta tese) buscou estimar a aparência física de três Neandertais, do homem de Denisova e de um Paleo-esquimó, além de predizer a aparência física de alguns indivíduos anônimos e não-anônimos da atualidade que possuem os seus respectivos genomas disponíveis. Nosso trabalho mostrou, dentre outras coisas, que os Neandertais podiam ter variados níveis de pigmentação, potencialmente na mesma escala observada nos humanos modernos. Uma repercussão bastante positiva na mídia nacional (Entrevista FAPESP – Apêndice IV) e internacional (Entrevista SCIENCENow – Apêndice V) ocorreu após esta publicação, pois nosso trabalho foi o primeiro a utilizar um conjunto grande de marcadores (124 SNPs) e também o primeiro a estimar a aparência física do homem de Denisova. Com este artigo foi possível também aferir a confiabilidade na predição de pigmentação através da análise de marcadores genéticos previamente selecionados, aspecto extremamente importante nas aplicações forenses de tais metodologias.

Posteriormente à publicação do artigo mencionado, um estudo publicado na prestigiosa revista internacional *Science* (Meyer *et al*., 2012) corroborou nossas inferências sobre a pigmentação do hominídeo de Denisova, indicando que o mesmo era portador de alelos que, nos humanos da atualidade, estão associados com pele escura, cabelos e olhos castanhos. Além de corroborar nossas inferências, Meyer e colegas também disponibilizaram no material suplementar de seu artigo, entre outras informações, uma tabela complementando genótipos faltantes no nosso estudo, deixando mais completa as informações genotípicas dos marcadores de pigmentação para o homem de Denisova. Para mais detalhes do atual estado da arte sobre genomas de hominíneos arcaicos, principalmente dados sobre o Denisova e Neanderthal, consulte a excelente revisão feita por Disotell (2012).

Sabemos que a abordagem utilizando o efeito aditivo, feita por nós (Cerqueira *et al*., 2012) é um tanto simplista, levando em consideração a complexidade envolvida nos mecanismos fisiológicos da pigmentação mencionados na Introdução desta tese, uma vez que consideramos apenas os genótipos homozigotos para os marcadores analisados, o que limitou nosso poder de predição. Entretanto, é possível trabalhar somente com as informações que temos disponíveis no momento, e outros trabalhos também utilizaram abordagens similares para uma infinidade de outras análises (Consultar Liu *et al*., 2010; Póspiech *et al*., 2011; Beleza *et al*., 2012). Esperamos que futuros trabalhos possam esclarecer o efeito dos heterozigotos no fenótipo da pigmentação para que estes possam ser testados também na predição de pigmentação usando uma metodologia similar àquela proposta por nós. Uma perspectiva que temos é aprimorar as análises estatísticas apresentadas (Cerqueira *et al*., 2012) a fim de testar o poder de predição com outros marcadores, utilizando abordagens e dados mais completos que estão sendo desenvolvidas pelo consórcio CANDELA. É possível antever que poderemos considerar o efeito dos heterozigotos no fenótipo de pigmentação e estimar o possível efeito de epistasia/interações gênicas nas análises computacionais. Com isto será possível selecionar um conjunto de marcadores específicos em populações miscigenadas para uso forense.

Os 18 SNPs envolvidos com pigmentação selecionados para o presente trabalho foram escolhidos baseados em dados funcionais e estudos de associação encontrados na literatura científica, conforme descrito na Introdução e nos artigos mencionados nos resultados desta tese. Um fato interessante nas análises realizadas com os 18 marcadores, é que alguns

marcadores clássicos (marcadores dos genes *MC1R* e *OCA2* aqui estudados, por exemplo) envolvidos com cor dos olhos e cabelos em Europeus, não foram consistentemente associados com estes fenótipos na nossa análise, mesmo que boa parte dos voluntários da nossa amostra possua ancestralidade predominantemente Européia (indivíduos do estado do Rio Grande do Sul). Para as análises da cor dos olhos e cabelos realizados no artigo descrito no Capítulo III desta tese, não realizamos análise estratificada para Gaúchos vs. Baianos como fizemos para o artigo descrito no Capítulo II, pois os olhos e cabelos foram inicialmente qualificados e algumas das categorias fenotípicas para estes fenótipos possuem pouco tamanho amostral para realizar tal estratificação, conforme é possível observar nas tabelas dos artigos mencionados.

Em Cerqueira *et al*. (2013; Capítulo II dos resultados desta tese) foi observado que alguns polimorfismos apresentaram associações diferentes entre Gaúchos e Baianos, com relação à cor da pele, o que pode estar relacionado aos diferentes perfis demográficos e história de mestiçagem destas populações. Estes dados são importantes sob vários aspectos: do ponto de vista de aplicação forense, para predizer características físicas de pigmentação em uma amostra, é necessário primeiro entender quais as variantes que diferem entre populações ou subpopulações, e então escolher àquelas que apresentam associações consistentes para uso prático, deixando àqueles marcadores com associações população-específica somente para uso restrito em alguns grupos geográficos. De acordo com Valenzuela *et al*. (2010), Spichenok *et al*. (2011) e Walsh *et al*. (2011, 2013), os ensaios para predição de aparência física podem ser desenvolvidos, independente da ancestralidade geográfica da população analisada. Nossos achados indicam que tal premissa deve ser considerada com restrições. Dos 18 marcadores, somente 3 se apresentaram consistentemente associados com cor da pele (rs1129038 - gene *HERC2*, rs1426654 - gene *SLC24A5*, e rs16891982 - gene *SLC45A2*) entre os dois grupos amostrais, enquanto outros marcadores se apresentaram associados somente em um ou outro grupo amostral. É conhecido também o fenômeno da convergência evolutiva, ou seja, a possibilidade de um mesmo fenótipo em populações distintas ter sido causado por diferentes mecanismos genéticos, devido a pressões seletivas similares (consultar Norton *et al*., 2007; e Beleza *et al*., 2012; para maiores detalhes).

A partir do modelo da dinâmica de história demográfica relatado no estudo de Beleza *et al*. (2012) é possível observar que as nossas indagações são plenamente plausíveis, no que diz respeito à recomendação de evitar extrapolar para várias populações resultados de associação

de genes vs. fenótipos de pigmentação. Os dados levantados para os 18 marcadores clássicos estudados na presente tese são muito importantes para mostrar esta questão em amostras miscigenadas.

Dos 18 SNPs estudados aqui, encontramos associação altamente significante para somente 3 marcadores em todas as três características de pigmentação investigadas (cor da pele, olhos e cabelos), que são: rs1129038 (gene *HERC2*), rs1426654 (gene *SLC24A5*) e rs16891982 (gene *SLC45A2*). Conforme também descrito nos nossos resultados, estes SNPs que estão consistentemente associados em mais de uma população nas nossas análises poderiam ser mantidos como polimorfismos preliminares realmente úteis para FDP, independentemente da ancestralidade geográfica da amostra analisada. Estudos adicionais em outras populações miscigenadas e não miscigenadas devem ser realizados para confirmar a consistência nestas associações. Uma importante recomendação do estudo do Ruiz *et al*. (2013) é a inclusão do SNP rs1129038 do gene *HERC2* no HIrisplex. Como dito, este SNP faz parte do conjunto de marcadores utilizados nesta tese e, a despeito do forte desequilíbrio de ligação entre este e o rs12913832 (Amos *et al*., 2011) (já utilizado no HIrisplex), o estudo de Ruiz *et al*. (2013) indicou a inclusão daquele por melhorar a preditabilidade para a cor do olho, o qual apresentou um efeito adicional ao rs12913832, depois de uma análise ajustada.

Em Cerqueira *et al*. (2013 – Capítulo III desta tese), também foi descrito o LD de alguns marcadores estudados. Este tipo de descrição é útil para o fim de desenvolvimento de polimorfismos "reservas" com o mesmo poder de predição e acurácia, no caso de alguns dos SNPs de um conjunto multiplex estabelecido não funcionar na genotipagem devido ao alto nível de degradação do DNA da amostra, por exemplo. Isto é útil em casos de desastres em massa e catástrofes naturais, onde é relativamente comum algumas amplificações não funcionarem. Portanto, uma vantagem de se ter SNPs reduntantes ou "SNPs extras" ao alcance da mão para análises de predição é a habilidade de lidar com perfis incompletos, nos quais não foi possível genotipar alguns marcadores (Ruiz *et al*., 2013).

Finalmente, esperamos que os achados da presente tese sirvam como ponto de partida para o desenvolvimento de conjunto de SNPs úteis para fenotipagem forense pelo DNA em populações miscigenadas. Esperamos também que nossos dados estimulem a discussão sobre o cuidado em algumas extrapolações sugeridas na literatura científica. Um esforço para estimular tal discussão e também para enriquecer a literatura com dados de populações

miscigenadas está sendo feito nos resumos apresentados em congressos no contexto da presente tese (Apêndice VI), bem como com as publicações que estão sendo preparadas com todos os países participantes do consórcio CANDELA (Apêndice VII). Concordamos com a sugestão de Kayser & Schneider (2009) de que o uso de marcadores geográficos devem ser encorajados para auxiliar a análise de predição de características físicas, no sentido de definir qual a provável população que o indivíduo analisado pertence, e então escolher o conjunto de marcadores de predição mais adequado em cada caso. Portanto, confirmar a possível origem genética geográfica da amostra preliminarmente ao uso de possíveis ensaios de predição de grupos populacionais específicos pode assegurar uma melhor confiabilidade nas estimativas da Fenotipagem Forense por DNA.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Alaluf S, Atkins D, Barrett K, Blount M, Carter N, Heath A. The impact of epidermal melanin on objective measurements of human skin colour. Pigment Cell Res. 15:119–126. 2002.

Amos CI, Wang LE, Lee JE, Gershenwald JE, Chen WV, Fang S, Kosoy R, Zhang M, Qureshi AA, Vattathil S, Schacherer CW, Gardner JM, Wang Y, Bishop DT, Barrett JH; GenoMEL Investigators, MacGregor S, Hayward NK, Martin NG, Duffy DL; Q-Mega Investigators, Mann GJ, Cust A, Hopper J; AMFS Investigators, Brown KM, Grimm EA, Xu Y, Han Y, Jing K, McHugh C, Laurie CC, Doheny KF, Pugh EW, Seldin MF, Han J, Wei Q. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. Hum Mol Genet. 20(24):5012-5023. 2011.

Aoki K. Sexual selection as a cause of human skin color variation: Darwin's hypothesis revisited. Ann Hum Biol. 29:589–608. 2002.

Badano JL & Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. Nat Rev Genet. 3:779–789. 2002.

Ballantyne J. Mass disaster genetics. Nat Genet. 15:329-331. 1997.

Ballaré CL, Caldwell MM, Flint SD, Robinson SA, Bornman JF. Effects of solar ultraviolet radiation on terrestrial ecosystems. Patterns, mechanisms, and interactions with climate change. Photochem Photobiol Sci. 10(2):226-241. 2011.

Barsh GS. The genetics of pigmentation: from fancy genes to complex traits. Trends Genet. 12:299–305. 1996.

Bastos JL, Dumith SC, Santos RV, Barros AJD, Del Duca GF, Gonçalves H, Nunes AP. Does the way I see you affect the way I see myself? Associations between interviewers' and interviewees' "color/race" in southern Brazil. Cad. Saúde Pública, Rio de Janeiro. 25(10):2111-2124. 2009.

Bauer M. RNA in forensic science. FSI Genetics. 1:69-74. 2007.

Beaumont KA, Shekar SN, Cook AL, Duffy DL, Sturm RA. Red hair is the null phenotype of MC1R. Hum Mutat. 29(8):E88-E94. 2008.

Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E, Shriver MD, Parra EJ, Rocha J. The timing of pigmentation lightening in Europeans. Mol Biol Evol. 30(1):24-35. 2012.

Bennett DC & Lamoreux ML. The color loci of mice: a genetic century. Pigment Cell Res. 16:333–344. 2003.

Biasutti R. Le razze e i popoli della terra. Torino: Unione Tipografico Editrice Torinese. 1953.

Birdsell JB. Microevolutionary patterns in aboriginal Australia. New York: Oxford University Press. 1993.

Bouakaze C, Keyser C, Crubézy E, Montagnon D, Ludes B. Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. Int J Legal Med. 123(4):315-325. 2009.

Bower C, Stanley F. Dietary folate as a risk factor for neural-tube defects: evidence from a case-control study in Western Australia. Med J Aust. 150(11):613–619. 1989.

Box N, Wyeth J, O'Gorman L, Martin N, Sturm R. Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. Hum Mol Genet. 6(11):1891–1897. 1997.

Branda R & Eaton J. Skin color and nutrient photolysis: an evolutionary hypothesis. Science. 201(4356):625–626. 1978.

Branicki W, Brudnik U, Kupiec T, Wolañska-Nowak P, Wojas-Pelc A: Determination of phenotype associated SNPs in the MC1R gene. J Forensic Sci. 52:349–354. 2007.

Branicki W, Brudnik U, Wojas-Pelc AL: Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. Ann Hum Genet. 73:160–170. 2009.

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

Buccimazza SS, Molteno CD, Dunne TT, Viljoen DL. Prevalence of neural tube defects in Cape Town, South Africa. Teratology. 50:194–199. 1994.

Budowle B & Baechtel FS. Modifications to improve the effectiveness of restriction fragment length polymorphism typing. Appl Theor Electrophor. 1:181-187. 1990.

Budowle B & Van Daal A. Forensically relevant SNP classes. Biotechniques. 44(5):603-608. 2008.

Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR-based short tandem repeat loci: law enforcement tools, p. 73–88. In Second European Symposium on Human Identification 1998, Promega Corporation, Madison, WI. 1998.

Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. J. Forensic Sci. 46:453-489. 2001.

Budowle B. SNP typing strategies. Forensic Sci Int. 146:S139-S142. 2004.

Bustamante J, Bredeston L, Malanga G, Mordoh J. Role of melanin as a scavenger of active oxygen species. Pigment Cell Res. 6:348–353. 1993.

Carter C. Ethnic origin and neural tube malformations. Developm Med Child eurol. 12:372–373. 1970.

Cavalli-Sforza L, Menozzi P, Piazza A. The history and geography of human genes. Princeton: Princeton University Press. p.266–267. 1994.

Cerqueira CCS, Amorim CEG, Salzano FM, Bortolini MC. Human pigmentation genes: forensic perspectives, general aspects and evolution. In: Forensic Science. ISBN: 978-1-61324-999-4. N. Yacine and R. Fellag, eds. © Nova Science Publishers, Inc. pp. 85-106. 2011.

Cerqueira CCS, Paixão-Côrtes VR, Zambra FMB, Salzano FM, Hunemeier T, Bortolini MC. Predicting Homo Pigmentation Phenotype Through Genomic Data: From Neanderthal to James Watson. Am J Hum Biol. 24(5):705-709. 2012.

Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of STR loci beyond human identification: implications for the development of new DNA typing systems. Electrophoresis. 20:1682-1696. 1999.

Chaplin G. Geographic distribution of environmental factors influencing human skin coloration. Am J Phys Anthropol. 125:292–302. 2004.

Cleaver JE & Crowley E. UV damage, DNA repair and skin carcinogenesis. Front Biosci. 7:1024–1043. 2002.

Clemens TL, Henderson SL, Adams JS, Holick MF. Increased skin pigment reduces the capacity of skin to synthesize vitamin D3. The Lancet. 1(8263):74–76. 1982.

Cook AL, Chen W, Thurber AE, Smit DJ, Smith AG, Bladen TG, Brown DL, Duffy DL, Pastorino L, Bianchi-Scarra G, Leonard JH, Stow JL, Sturm RA. Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. J Invest Dermatol. 129(2):392-405. 2009.

Coplan RJ, Coleman B, Rubin KH. Shyness and little boy blue: iris pigmentation, gender, and social wariness in preschoolers. Dev Psychobiol. 32:37–44. 1998.

Costin GE & Hearing VJ. Human skin pigmentation: melanocytes modulate skin color in response to stress. Faseb J. 21:976–994. 2007.

Costin GE, Valencia JC, Wakamatsu K, Ito S, Solano F, Milac AL, Vieira WD, Yamaguchi Y, Rouzaud F, Petrescu AJ, Lamoreux ML, Hearing VJ. Mutations in dopachrome tautomerase (Dct) affect eumelanin/pheomelanin synthesis, but do not affect intracellular trafficking of the mutant protein. Biochem J. 391:249–259. 2005.

Daniels FJr. Man and radiant energy: solar radiation. In (D. B. Dill & E. F. Adolph, Eds) Adaptation to the Environment. Handbook of Physiology, pp. 969–985. Washington, D.C.: American Physiological Society. 1964.

Darwin C. The descent of man. Murray: London. 1871.

Davenport GC & Davenport CB. Heredity of eye-color in man. Science. 26:590–592. 1907.

Diamond J. The rise and fall of the third chimpanzee. London: Radius. 1991.

Dimisianos G, Stefanaki I, Nicolaou V, Sypsa V, Antoniou C, Poulou M, Papadopoulos O, Gogas H, Kanavakis E, Nicolaidou E, Katsambas AD, Stratigos AJ. A study of a single variant allele (rs1426654) of the pigmentation-related gene SLC24A5 in Greek subjects. Exp Dermatol. 18(2):175-177. 2009.

Disotell TR. Archaic human genomics. Am J Phys Anthropol. 149(Suppl 55):24-39. 2012.

Donaldson DD. The significance of the iris in mongolids: Brushfield spots. Arch. Ophthalmol. 65:26–31. 1961.

Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Barta C, Lu RB, Zhukova OV, Kim JJ, Siniscalco M, New M, Li H, Kajuna SL, Manolopoulos VG, Speed WC, Pakstis AJ, Kidd JR, Kidd KK. A global view of the OCA2-HERC2 region and pigmentation. Hum Genet. 131(5):683-696. 2012.

Draus-Barini J, Walsh S, Pośpiech E, Kupiec T, Głąb H, Branicki W, Kayser M. Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. Investig Genet. 4(1):3. 2013.

Duffy DL, Box NF, Chen W, Palmer JS, Montgomery GW, James MR, Hayward NK, Martin NG, Sturm RA. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. Hum Mol Genet. 13(4):447-461. 2004.

Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, Hayward NK, Martin NG, Sturm RA. A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am J Hum Genet. 80:241–252. 2007.

Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet. 49:746-756. 1991.

Edwards M, Bigham A, Tan J, Li S, Gozdzik A, Ross K, Jin L, Parra EJ. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations:

Further evidence of convergent evolution of skin pigmentation. PLoS Genet 6:e1000867. 2010.

Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L: Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet. 123:177–187. 2008.

Eiberg H & Mohr J. Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. Eur J Hum Genet. 4:237–241. 1996.

Eiberg H & Mohr J. Major locus for red hair color linked to MNS blood groups on chromosome 4. Clin Genet. 32:125–128. 1987.

Elwood JM & Elwood JH. Epidemiologyof Anencephalus and Spina Bifida. Oxford: Oxford University Press. 1980.

Fan S, Dyer CR, Hubbard L. Quantification and correction of iris color. Technical Report 1495, Department of Computer Sciences, University of Wisconsin-Madison. 2003.

Fernstrom JD & Fernstrom MH. Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain. J Nutr. 137(6 Suppl 1): 1539S-1547S. 2007.

Fitzpatrick TB, Eisen AZ, Wolff K, Freedberg IM, Austen KF. In Dermatology in General Medicine. Third Edition. (McGraw-Hill, New York). 1987.

Fogelman Y, Rakover Y, Luboshitsky R. High prevalence of vitamin D deficiency among Ethiopian women immigrants to Israel: exacerbation during pregnancy and lactation. Isr J Med Sci. 31:221–224. 1995.

Frandberg P, Doufexis M, Kapas S, Chhajlani J. Human pigmentation phenotype: a point mutation generates nonfunctional MSH receptor. Biochem Biophys Res Commun. 245:490–492. 1998.

Frost P. European hair and eye color: a case of frequency dependent sexual selection? Evol Hum Behav. 27:85–103. 2006.

Frudakis T, Terravainen T, Thomas M. Multilocus OCA2 genotypes specify human iris colors. Hum Genet. 122(3-4):311-326. 2007.

Frudakis TN. Molecular Photofitting – Predicting Ancestry and Phenotype Using DNA (Burlington, MA, USA: Academic Press). 2008.

Frudakis T, Thomas M, Gaskin Z, Venkateswarlu K, Chandra KS, Ginjupalli S, Gunturi S, Natrajan S, Ponnuswamy VK, Ponnuswamy KN. Sequences associated with human iris pigmentation. Genetics. 165(4):2071–2083. 2003.

Gao X, Simon KC, Han J, Schwarzschild MA, Ascherio A. Genetic determinants of hair color and Parkinson's disease risk. Ann Neurol. 65 (1): 76-82. 2009.

Gehring MP. SNPs e pigmentação humana – informação forense. Monografia de especialização apresentada à Pós-Graduação *Latu Sensu* em Biologia e Genética Forense da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Orientadora: Dra. Clarice Sampaio Alho. 2010.

German EJ, Hurst MA, Wood D, Gilchrist J. A novel system for the objective classification of iris colour and its correlation with response to 1% tropicamide. Ophthal Physiol Opt. 18:103–110. 1998.

Gill P, Werrett DJ, Budowle B, Guerrieri R. An assessment of whether SNPs will replace STRs in national DNA databases-joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Sci Justice. 44:51-53. 2004.

Giusti AM & Budowle B. Chemiluminescence-based detection system for human DNA quantitation and restriction fragment length polymorphism (RFLP) analysis. Appl Theor Electrophor. 5:89-98. 1995.

Gonçalves VF, Carvalho CM, Bortolini MC, Bydlowski SP, Pena SD. The phylogeography of African Brazilians. Hum Hered. 65(1):23-32. 2008.

Goodwin W, Linacre A, Hadi S. An introduction to forensic genetics. 2nd. edition. Wiley Blackwell. ISBN 978-0-470-71019-7. 2011.

Graf J, Voisey J, Hughes I, van Daal, A. Promoter polymorphisms in the MATP (SLC45A2) gene are associated with normal human skin color variation. Hum Mut. 28:710-717. 2007.

Graf J, Hodgson R, van Daal A. Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum Mutat. 25:278–284. 2005.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. A draft sequence of the Neandertal genome. Science. 328(5979):710-722. 2010.

Grimes EA, Noake PJ, Dixon L, Urquhart A. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. Forensic Sci Int. 122(2-3):124–129. 2001.

Guerreiro-Junior V, Bisso-Machado R, Marrero A, Hünemeier T, Salzano FM, Bortolini MC. Genetic signatures of parental contribution in black and white populations in Brazil. Genet Mol Biol. 32(1):1-11. 2009.

Halliday GM. Inflammation, genemutation and photoimmunosuppression in response to UVR-induced oxidative damage contributes to photocarcinogenesis. Mutat Res. 571:107–120. 2005.

Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL. Evidence for variable selective pressures at MC1R. Am J Hum Genet. 66(4):1351-1361. 2000.

Healy E, Flannagan N, Ray A, Todd C, Jackson IJ, Matthews JN, Birch-Machin MA, Rees JL. Melanocortin-1-receptor gene and sun sensitivity in individuals without red hair. Lancet. 355(9209):1072-1073. 2000.

Henderson JB, Dunnigan MG, McIntosh WB, Abdul-Motaal AA, Gettinby G & Glekin BM. The importance of limited exposure to ultraviolet radiation and dietary factors in the aetiology of Asian rickets: a risk factor model. Q J Med. 63:413–425. 1987.

Hill HZ & Hill GJ. UVA, pheomelanin and the carcinogenesis of melanoma. Pigment Cell Res. 13(8 Suppl):140–144. 2000.

Holick MF. The vitamin D epidemic and its health consequences. J Nutr. 135:2739S–2748S. 2005.

Holick MF. Vitamin D deficiency. N Engl J Med. 357:266–281. 2007.

Holick MF. Vitamin D: a millennium perspective. J Cell Biochem. 88:296–307. 2003.

Holick MF. Vitamin D: importance in the prevention of cancers, type 1 diabetes, heart disease and osteoporosis. Am J Clin Nutr. 79:362–371. 2004.

Holick MF, MacLaughlin JA, Doppelt SH. Regulation of cutaneous previtamin D3 photosynthesis in man: skin pigment is not an essential regulator. Science. 211:590–593. 1981.

Holmes SJ & Loomis HM. The heredity of eye color and hair color in man. Biol Bull. 18:50–65. 1909.

Hünemeier T, Carvalho C, Marrero AR, Salzano FM, Junho Pena SD, Bortolini MC. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. Am J Phys Anthropol. 133(2):854-867. 2007.

Hurst CC. On the inheritance of eye-colour in man. Proc R Soc Med Lond B Biol Sci. 80:85–96. 1908

Iida R, Ueki M, Takeshita H, Fujihara J, Nakajima T, Kominato Y, Nagao M, Yasuda T. Genotyping of five single nucleotide polymorphisms in the OCA2 and HERC2 genes associated with blue-brown eye color in the Japanese population. Cell Biochem Funct. 27:323–327. 2009.

Imesch PD, Wallow IH, Albert DM. The color of the human eye: a review of morphologic correlates and of some conditions that affect iridial pigmentation. Surv Ophthalmol. 41(Suppl 2):S117-S123. 1997.

Ito S. The IFPCS presidential lecture: a chemist's view of melanogenesis. Pigment Cell Res. 16: 230–236. 2003.

Izagirre N, García I, Junquera C, de la Rúa C, Alonso S. A scan for signatures of positive selection in candidate loci for skin pigmentation in humans. Mol Biol Evol. 23(9):1697-1706. 2006.

Jablonski NG & Chaplin G. Human skin pigmentation as an adaptation to UV radiation. Proc Natl Acad Sci USA. 107(Suppl 2):8962-8968. 2010.

Jablonski NG & Chaplin G. Human skin pigmentation, migration and disease susceptibility. Philos Trans R Soc Lond B Biol Sci. 367(1590):785-792. 2012.

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

Jablonski NG. The evolution of human skin colouration and its relevance to health in the modern world. J R Coll Physicians Edinb. 42(1):58-63. 2012.

Jackson IJ. Homologous pigmentation mutations in human, mouse and other model organisms. Hum Mol Genet. 6:1613–1624. 1997.

Jeffreys AJ, Wilson V, Thein SL. Hypervariable minisatellite regions in human DNA. Nature. 314:67-73. 1985.

Jobim LF, Costa LR, Silva M. Identificação humana: identificação pelo DNA, Identificação Médico-legal, e perícias odontológicas. (Tratado de perícias criminalísticas, Org. Domingos Tocchetto). Volume II. Millennium editora. Campinas/SP. ISBN 85.7625-064-0. 2006.

Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR. A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet. 70(3):770-775. 2002.

Kayser M & Knijff P: Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet. 12:179–192. 2011.

Kayser M, Liu F, Janssens AC, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van Ijcken WF, den Dunnen JT, Heath S, Zelenika D, Despriet DD, Klaver CC, Vingerling JR, de Jong PT, Hofman A, Aulchenko YS, Uitterlinden AG, Oostra BA, van Duijn CM: Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet. 82(2):411–423. 2008

Kayser M & Schneider PM. DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. Forensic Sci Int Genet. 3:154–161. 2009.

Khan R, Khan BS. Diet, disease and pigment variation in humans. Med Hypotheses. 75 (4): 363-367. 2010.

King R, Hearing V, Creek D, Oetting W, Albinism. In: Scriver C, Beaudet A, Sly W, Valle D, editors. The metabolic and molecular basis of inherited diseases. 7th ed. New York: McGraw Hill. 1994.

Kobayashi H & Kohshima S. Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. J. Hum. Evol. 40:419–435. 2001.

Köstner K, Denzer N, Müller CS, Klein R, Tilgen W, Reichrath J. The relevance of vitamin D receptor (VDR) gene polymorphisms for cancer: a review of the literature. Anticancer Res. 29(9):3511-3536. 2009.

Krol ES & Liebler DC. Photoprotective actions of natural and synthetic melanins. Chem Res Toxicol. 11:1434–1440. 1998.

Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, Hughes D, Rohland N, Pilli E, Longo L, Condemi S, de la Rasilla M, Fortea J, Rosas A, Stoneking M, Schöneberg T, Bertranpetit J, Hofreiter M. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. Science. 318(5855):1453-1455. 2007.

Lamparelli RD, Bothwell TH, MacPhail AP, van der Westhuyzen J, Baynes RD, MacFarlane BJ. Nutritional anaemia in pregnant coloured women in Johannesburg. S Afr Med J. 73(8):477-481. 1988.

Lao O, De Gruijter JM, Van Duijn K, Navarro A, Kayser M. Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. Ann Hum Genet. 71:354–369. 2007.

Larsson M, Pedersen NL, Stattin H. Importance of genetic effects for characteristics of the human iris. Twin Res. 6(3):192-200. 2003.

Larsson M. Human Iris Characteristics as Biomarkers for Personality. Örebro studies in psychology. ISSN 1651-1328. Örebro University, Department of Behavioural, Social and Legal Sciences. Sweden. 2007.

Latreille J, Ezzedine K, Elfakir A, Ambroisine L, Gardinier S, Galan P, Hercberg S, Gruber F, Rees J, Tschachler E, Guinot C. MC1R gene polymorphism affects skin color and phenotypic features related to sun sensitivity in a population of French adult women. Photochem Photobiol. 85(6):1451-1458. 2009.

Lawrence VA. Demographic analysis of serum folate and folate-binding capacity in hospitalized patients. Acta Haematol. 69(5):289-293. 1983.

Lee MJ & Stephenson DA. Recent developments in neurofibromatosis type 1. Curr Opin Neurol. 20:135–341. 2007.

Lin JY & Fisher DE. Melanocyte biology and skin pigmentation. Nature. 445:843–50. 2007.

Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens AC, Kayser M. Eye color and the prediction of complex phenotypes from genotypes. Curr Biol. 19(5):R192-R193. 2009.

Liu F, Wollstein A, Hysi PG, Ankra-Badu GA, Spector TD, Park D, Zhu G, Larsson M, Duffy DL, Montgomery GW, Mackey DA, Walsh S, Lao O, Hofman A, Rivadeneira F, Vingerling JR, Uitterlinden AG, Martin NG, Hammond CJ, Kayser M. Digital quantification of human eye color highlights genetic association of three new loci. PLoS Genet. 6(5):e1000934. 2010.

Loomis WF. Skin pigment regulation of vitamin-D biosynthesis in man. Science. 157(788):501–506. 1967.

Lucock M. Folic acid: nutritional biochemistry, molecular biology, and role in disease processes. Mol Genet Metab. 71:121–138. 2000.

Mackey DA, Wilkinson CH, Kearns LS, Hewitt AW. Classification of iris colour: review and refinement of a classification schema. Clin Experiment Ophthalmol. 39(5):462-471. 2011.

Mackintosh J. The antimicrobial properties of melanocytes, melanosomes, and melanin and the evolution of black skin. J Theor Biol. 211(2):101–113. 2001.

Makova K & Norton H. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. Peptides. 26:1901–1908. 2005.

Marien P, Brouns R, Engelborghs S, Wackenier P, Verhoeven J, Ceulemans B, De Deyn PP. Cerebellar cognitive affective syndrome without global mental retardation in two relatives with Gillespie syndrome. Cortex. 44:54–67. 2008.

Marrero AR, Bravi C, Stuart S, Long JC, Pereira das Neves Leite F, Kommers T, Carvalho CM, Pena SD, Ruiz-Linares A, Salzano FM, Bortolini MC. Pre- and post-Columbian gene and cultural continuity: the case of the Gaucho from southern Brazil. Hum Hered. 64(3):160-171. 2007.

Marrero AR, Das Neves Leite FP, De Almeida Carvalho B, Peres LM, Kommers TC, Da Cruz IM, Salzano FM, Ruiz-Linares A, Da Silva Júnior WA, Bortolini MC. Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. Am J Hum Biol. 17(4):496-506. 2005.

Mathur U, Datta SL, Mathur BB. The effect of aminopterin induced folic acid deficiency on spermatogenesis. Fertil Steril. 28(12):1356–1360. 1977.

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

Melgosa M, Rivas MJ, Gomez L, Lita E. Towards a colorimetric characterization of the human iris. Ophthal Physiol Opt. 20:252–260. 2000.

Mengel-From J, Wong TH, Morling N, Rees JL, Jackson IJ: Genetic determinants of hair and eye colours in the Scottish and Danish populations. BMC Genet. 10:88. 2009.

Mengel-From J, Børsting C, Sanchez JJ, Eiberg H, Morling N. Human eye colour and HERC2, OCA2 and MATP. FSI Genet. 4(5):323-328. 2010.

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andrés AM, Eichler EE, Slatkin M, Reich D, Kelso J, Pääbo S. A high-coverage genome sequence from an archaic Denisovan individual. Science. 338(6104):222-226. 2012.

Myles S, Somel M, Tang K, Kelso J, Stoneking M. Identifying genes underlying skin pigmentation differences among human populations. Hum Genet. 120:613–621. 2007.

Neitzke-Montinelli V, Urmenyi TP, Rondinelli E, Cabello PH, Silva R, Moura-Neto RS. Polymorphisms upstream of the melanocortin-1 receptor coding region are associated with human pigmentation variation in a Brazilian population. Am J Hum Biol. 24(6):853-855. 2012.

Nielsen KP, Zhao L, Stamnes JJ, Stamnes K, Moan J. The importance of the depth distribution of melanin in skin for DNA protection and other photobiological processes. J Photochem Photobiol B. 82(3):194-198. 2006.

Niggemann B, Weinbauer G, Vogel F, Korte R. A standardized approach for iris color determination. Int J Toxicol. 22:49–51. 2003.

Norman AW. From vitamin D to hormone D: Fundamentals of the vitamin D endocrine system essential for good health. Am J Clin Nutr. 88:491S–499S. 2008.

Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol. 24:710–722. 2007.

Oetteking B. Rudolf Martin. Am Anthropol. 28:414–417. 1926.

Off MK, Steindal AE, Porojnicu AC, Juzeniene A, Vorobey A, Johnsson A, Moan J. Ultraviolet degradation of folic acid. J Photochem Photobiol B. 80:47–55. 2005.

Ortonne JP & Prota G. Hair melanins and hair color: ultrastructural and biochemical aspects. J Invest Dermatol. 101(1 Suppl):82S–89S. 1993.

Pakstis AJ, Speed WC, Fang R, Hyland FC, Furtado MR, Kidd JR, Kidd KK. SNPs for a universal individual identification panel. Hum Genet. 127(3):315-324. 2010.

Palmer JS, Duffy DL, Box NF, Aitken JF, O'Gorman LE, Green AC, Hayward NK, Martin NG, Sturm RA. Melanocortin-1 receptor polymorphisms and risk of melanoma: is the association explained solely by pigmentation phenotype? Am J Hum Genet. 66(1):176-186. 2000.

Pandita KK, Razdan S, Kudyar RP, Beigh A, Kuchay S, Banday T. "Excess gooD can be Dangerous". A case series of iatrogenic symptomatic hypercalcemia due to hypervitaminosis D. Clin Cases Miner Bone Metab. 9(2):118-120. 2012.

Pandolf KB, Gange RW, Latzka WA, Blank IH, Kraning KK 2nd, Gonzalez RR. Human thermoregulatory responses during heat exposure after artificially induced sunburn. Am. J. Physiol. 262:R610–R616. 1992.

Parra EJ. Human pigmentation variation: evolution, genetic basis, and implications for public health. Am J Phys Anthropol. 45:85-105. 2007.

Parra EJ, Kittles RA, Shriver MD. Implications of correlations between skin color and genetic ancestry for biomedical research. Nat Genet. 36(11 Suppl):S54-60. 2004.

Passeron T, Valencia JC, Bertolotto C, Hoashi T, Le Pape E, Takahashi K, Ballotti R, Hearing VJ. SOX9 is a key player in ultraviolet B-induced melanocyte differentiation and pigmentation. Proc Natl Acad Sci USA. 104:13984–13989. 2007.

Pena SDJ, Prado VF, Epplen JT. DNA diagnosis of human genetic individuality. J Mol Med. 73:555-564. 1995.

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, Pritchard JK. Signals of recent positive selection in a worldwide sample of human populations. Genome Res.19(5):826-837. 2009.

Pośpiech E, Draus-Barini J, Kupiec T, Wojas-Pelc A, Branicki W. Gene-gene interactions contribute to eye colour variation in humans. J Hum Genet. 56(6):447-455. 2011.

Post PW, Daniels FJr, Binford RTJr. Cold injury and the evolution of “white”skin. Hum Biol. 47(1):65–80. 1975.

Posthuma D, Visscher PM, Willemsen G, Zhu G, Martin NG, Slagboom PE, de Geus EJ, Boomsma DI. Replicated linkage for eye color on 15q using comparative ratings of sibling pairs. Behav Genet. 36(1):12-17. 2006.

Przeworski M, Hudson R, DiRienzo A. Adjusting the focus on human variation. Trends Genet. 16(7):296–302. 2000.

Quillen EE & Shriver MD. Unpacking human evolution to find the genetic determinants of human skin pigmentation. J Invest Dermatol. 131(E1):E5-E7. 2011.

Rana BK, Hewett-Emmett D, Jin L, Chang BH, Sambuughin N, Lin M, Watkins S, Bamshad M, Jorde LB, Ramsay M, Jenkins T, Li WH. High polymorphism at the human melanocortin 1 receptor locus. Genetics. 151(4):1547-1557. 1999.

Rebbeck TR, Kanetsky PA, Walker AH, Holmes R, Halpern AC, Schuchter LM, Elder DE, Guerry D. P gene as an inherited biomarker of human eye color. Cancer Epidemiol Biomarkers Prev. 11:782–784. 2002.

Rees JL. Genetics of hair and skin colour. Annu Rev Genet. 37:67–90. 2003.

Rees JL. The genetics of sun sensitivity in humans. Am J Hum Genet. 75:739-751. 2004.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo S.Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 468(7327):1053-60. 2010.

Relethford JH. Hemispheric difference in human skin color. Am J Phys Anthropol. 104:449–457. 1997.

Robins AH. Biological perspectives on human pigmentation. Cambridge, NY: Cambridge University Press. 1991.

Rohlfs RV, Fullerton SM, Weir BS. Familial identification: population structure and relationship distinguishability. PLoS Genet. 8(2):e1002469. 2012.

Rowe PJ & Evans P. Ball color, eye color, and a reactive motor skill. Percept Mot Skills. 79:671–674. 1994.

Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares de Cal M, Cruz R, Maroñas O, Söchtig J, Fondevila M, Rodriguez-Cid MJ, Carracedo A, Lareu MV. Further development of forensic eye color predictive tests. FSI Genet. 7(1):28-40. 2013.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, International HapMap Consortium. Genome-wide detection and characterization of positive selection in human populations. Nature. 449(7164):913-918. 2007.

Samokhvalov A, Hong L, Liu Y, Garguilo J, Nemanich RJ, Edwards GS, Simon JD. Oxidation potentials of human eumelanosomes and pheomelanosomes. Photohem Photobiol. 81:145–148. 2005.

Schioth HB, Phillips SR, Rudzish R, Birch-Machin MA, Wikberg JE, Rees JL. Loss of function mutations of the human melanocortin 1 receptor are common and are associated with red hair. Biochem Biophys Res Commun. 260:488–491. 1999.

Schneider PM. Beyond STRs: The Role of Diallelic Markers in Forensic Genetics. Transfus Med Hemother. 39(3):176-180. 2012.

Shaw GM, Jensvold NG, Wasserman CR, Lammer EJ. Epidemiologic characteristics of phenotypically distinct neural tube defects among 0·7 million California births, 1983–1987. Teratology. 49:143–149. 1994.

Shriver M & Parra E. Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. Am J Phys Anthropol. 112(1):17–27. 2000.

Slominski A, Wortsman J, Plonka PM, Schallreuter KU, Paus R, Tobin DJ. Hair follicle pigmentation. J Invest Dermatol 124:13–21. 2005.

Smith R, Healy E, Siddiqui S, Flanagan N, Steijlen PM, Rosdahl I, Jacques JP, Rogers S, Turner R, Jackson IJ, Birch-Machin MA, Rees JL. Melanocortin 1 receptor variants in an Irish population. J Invest Dermatol. 111(1):119-122. 1998.

Soejima M, Tachida H, Ishida T, Sano A, Koda Y. Evidence for recent positive selection at the human AIM1 locus in a European population. Mol Biol Evol. 23:179–188. 2006.

Souza RG. Aplicação forense de SNPs informativos de fenótipo relacionados à pigmentação da íris humana. Monografia de especialização apresentada à Pós-Graduação *Latu Sensu* em Biologia e Genética Forense da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Orientadora: Dra. Bianca de Almeida Carvalho. 2012.

Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, Caragine T, Prinz M, Wurmbach E: Prediction of eye and skin color in diverse populations using seven SNPs. FSI Genet. 5:472–478. 2011.

Spinney L. Eyewitness identification: line-ups on trial. Nature. 453(7194):442-444. 2008.

Sterling-Levis K & Williams K. What is the connection between red hair and Tourette syndrome? Med Hypotheses. 73 (5):849-853. 2009.

Steingrimsson E, Copeland NG, Jenkins NA. Melanocytes and the microphthalmia transcription factor network. Annu Rev Genet. 38:365–411. 2004.

Stjernschantz JW, Albert DM, Hu DN, Drago F, Wistrand PJ. Mechanism and clinical significance of prostaglandin-induced iris pigmentation. Surv Ophthalmol. 47(Suppl 1):S162-S175. 2002.

Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, Ginger RS, Green MR, van der Ouderaa FJ, Cox DR. A genomewide association study of skin pigmentation in a South Asian population. Am J Hum Genet. 81(6):1119-1132. 2007.

Sturm RA & Duffy DL. Human pigmentation genes under environmental selection. Genome Biol. 13(9):248. 2012.

Sturm RA & Frudakis TN. Eye colour: portals into pigmentation genes and ancestry. Trends Genet. 20(8):327-332. 2004.

Sturm RA & Larsson M. Genetics of human iris colour and patterns. Pigment Cell Melanoma Res. 22(5):544-562. 2009.

Sturm RA, Box NF, Ramsay M. Human pigmentation genetics: the difference is only skin deep. Bioessays. 20(9):712-721. 1998.

Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, Martin NG, Montgomery GW: A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet. 82:424–431. 2008.

Sturm RA, Teasdale RD, Box NF. Human pigmentation genes: identification, structure and consequences of polymorphic variation. Gene. 277(1-2):49-62. 2001.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G, Pálsson S, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Vermeulen SH, Goldstein AM, Tucker MA, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K: Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet. 40:835–837. 2008

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, Jakobsdottir M, Steinberg S, Pálsson S, Jonasson F, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Aben KK, Kiemeney LA, Olafsson JH, Gulcher J, Kong A, Thorsteinsdottir U, Stefansson K. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet. 39(12):1443-1452. 2007.

Tadokoro T, Rouzaud F, Itami S, Hearing VJ, Yoshikawa K. The inhibitory effect of androgen and sex-hormone-binding globulin on the intracellular cAMP level and tyrosinase activity of normal human melanocytes. Pigment Cell Res. 16:190–197. 2003.

Tadokoro T, Yamaguchi Y, Batzer J, Coelho SG, Zmudzka BZ, Miller SA, Wolber R, Beer JZ, Hearing VJ. Mechanisms of skin tanning in different racial/ethnic groups in response to ultraviolet radiation. J Invest Dermatol. 124:1326–1332. 2005.

Takeuchi S, Zhang W, Wakamatsu K, Ito S, Hearing VJ, Kraemer KH, Brash DE. Melanin acts as a potent UVB photosensitizer to cause an atypical mode of cell death in murine skin. Proc Natl Acad Sci USA. 101:15706–15801. 2004.

The 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 467(7319):1061-1073. 2010.

Ticho BH, Hilchie-Schmidt C, Egel RT, Traboulsi EI, Howarth RJ, Robinson D. Ocular findings in Gillespie-like syndrome: association with a new PAX6 mutation. Ophthalmic Genet. 27:145–149. 2006.

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

Ullrich SE. Mechanisms underlying UV-induced immune suppression. Mutat Res. 571:185–205. 2005.

Valenzuela RK, Henderson MS, Walsh MH, Garrison NA, Kelch JT, Cohen-Barak O, Erickson DT, John Meaney F, Bruce Walsh J, Cheng KC, Ito S, Wakamatsu K, Frudakis T, Thomas M, Brilliant MH. Predicting phenotype from genotype: normal pigmentation. J Forensic Sci. 55(2):315-322. 2010.

Valverde P, Healy E, Jackson I, Rees J, Thody AJ. Variants of the melanocyte stimulating hormone receptor gene are associated with red hair and fair skin. Nat Genet. 11(3): 328–30. 1995.

Visser M, Kayser M, Palstra RJ. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res. 22(3):446-455. 2012.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 4:e72. 2006.

Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, Kayser M, Ballantyne KN: Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. FSI Genet. 5:464–471. 2011a.

Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M: IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. FSI Genet. 5:170–180. 2011b.

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

Walsh S, Wollstein A, Liu F, Chakravarthy U, Rahu M, Seland JH, Soubrane G, Tomazzoli L, Topouzis F, Vingerling JR, Vioque J, Fletcher AE, Ballantyne KN, Kayser M. DNA-based eye colour prediction across Europe with the IrisPlex system. FSI Genet. 6(3): 330-340. 2012.

Werrett DJ. The new frontiers of forensics. Polic Today. 11: 29–31. 2005.

Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, Gill P. Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. BioTechniques 18:670-677. 1995.

Wielgus AR & Sarna T. Melanin in human irides of different color and age of donors. Pigment Cell Res 18:454–464. 2005.

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. Localizing recent adaptive evolution in the human genome. PLoS Genet. 3:e90. 2007.

Wiswell TE, Tuttle DJ, Northam RS, Simonds GR. Major congenital neurologic malformations. A 17-year survey. Am J Dis Child. 144: 61–67. 1990.

Wright S. Color inheritance in mammals. J Hered. 9: 227–240. 1918.

Wyman AR & White R. A highly polymorphic locus in human DNA. Proc Natl Acad Sci USA 77:6754-6758. 1980.

Xu X, Thornwall M, Lundin L, Chhajlani V. Val92Met variant of the melanocyte stimulating hormone receptor gene. Nat Genet 14:384. 1996.

Yakubu A & Mabogunje OA. Skin cancer in African albinos. Acta Oncol 32:621–622. 1993.

Yamaguchi Y & Hearing VJ. Physiological factors that regulate skin pigmentation. Biofactors. 35(2):193-199. 2009.

Ye T, Hong L, Garguilo J, Pawlak A, Edwards GS, Nemanich RJ, Sarna T, Simon JD. Photoionization thresholds of melanins obtained from free electron laser-photoelectron emission microscopy, femtosecond transient absorption spectroscopy and electron paramagnetic resonance measurements of oxygen photoconsumption. Photochem Photobiol 82:733–737. 2006.

Zhu G, Evans DM, Duffy DL, Montgomery GW, Medland SE, Gillespie NA, Ewen KR, Jewell M, Liew YW, Hayward NK, Sturm RA, Trent JM, Martin NG. A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q. Twin Res 7:197–210. 2004.

6. ANEXOS

6.1. ANEXO I

(Aprovação do projeto de pesquisa pela Universidade Federal do Rio Grande do Sul - UFRGS)

U F R G S
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

PRÓ-REITORIA DE PESQUISA

Comitê De Ética Em Pesquisa Da Ufrgs

PROPESQ

CARTA DE APROVAÇÃO

Comitê De Ética Em Pesquisa Da Ufrgs analisou o projeto:

Número: 18208
Título: FENOTIPO, ANCESTRALIDADE GENOMICA E DINAMICA DE MESTICAGEM

Pesquisadores:

Equipe UFRGS:

MARIA CATIRA BORTOLINI DA SILVA - coordenador desde 01/07/2010
FRANCISCO MAURO SALZANO - pesquisador desde 01/07/2010
Tábita Hünemeier - pesquisador desde 01/07/2010
Caio Cesar Silva de Cerqueira - pesquisador desde 01/07/2010

Equipe Externa:

Andres Ruiz Linares - coordenador desde 01/07/2010
Jorge Gomez Valdes - pesquisador desde 01/07/2010
Victor Acuna Alonzo - pesquisador desde 01/07/2010

Comitê De Ética Em Pesquisa Da Ufrgs aprovou o mesmo, em reunião realizada em 03/09/2010 - Sala de Reuniões do Gabinete do Reitor (Ex Salão Vermelho) - Prédio Reitoria, 6° andar, por estar adequado ética e metodologicamente e de acordo com a Resolução 196/96 e complementares do Conselho Nacional de Saúde.

Porto Alegre, Segunda-Feira, 13 de Setembro de 2010

JOSE ARTUR BOGO CHIES
Coordenador da comissão de ética

6.2. ANEXO II

(Aprovação do projeto de pesquisa pelo Hospital de Clínicas de Porto Algre - HCPA)

HCPA - HOSPITAL DE CLÍNICAS DE PORTO ALEGRE
GRUPO DE PESQUISA E PÓS-GRADUAÇÃO

COMISSÃO CIENTÍFICA E COMISSÃO DE PESQUISA E ÉTICA EM SAÚDE

A Comissão Científica e a Comissão de Pesquisa e Ética em Saúde, que é reconhecida pela Comissão Nacional de Ética em Pesquisa (CONEP)/MS como Comitê de Ética em Pesquisa do HCPA e pelo Office For Human Research Protections (OHRP)/USDHHS, como Institutional Review Board (IRB00000921) analisaram o projeto:

Projeto: 100565 **Versão do Projeto:** 11/01/2011 **Versão do TCLE:** 11/01/2011

Pesquisadores:
LAVINIA SCHULER
DÂNAE LONGO
TABITA HUNEMEIER
CAIO CESAR SILVA DE CERQUEIRA
MARIA CÁTIRA BORTOLINI

Título: Fenótipo, ancestralidade genética e dinâmica de miscigenação no Brasil

- Este projeto foi Aprovado em seus aspectos éticos e metodológicos, inclusive quanto ao seu Termo de Consentimento Livre e Esclarecido, de acordo com as Diretrizes e Normas Internacionais e Nacionais, especialmente as Resoluções 196/96 e complementares do Conselho Nacional de Saúde. Os membros do CEP/HCPA não participaram do processo de avaliação dos projetos onde constam como pesquisadores. Toda e qualquer alteração do Projeto, assim como os eventos adversos graves, deverão ser comunicados imediatamente ao CEP/HCPA. Somente poderão ser utilizados os Termos de Consentimento onde conste a aprovação do GPPG/HCPA.
- De acordo com a regulamentação da Resolução 340/2004 do CNS/MS o CEP/HCPA foi credenciado, através da Carta Circular Nº 037 CONEP/CNS/MS de 11 de agosto de 2004, para dar aprovação final para este projeto.

Porto Alegre, 18 de janeiro de 2011.

Profª Nadine Clausell
Coordenadora GPPG e CEP/HCPA

6.3. ANEXO III

(Aprovação do projeto de pesquisa pela Universidade Estadual do Sudoeste da Bahia - UESB)

Universidade Estadual do Sudoeste da Bahia – UESB
Autorizada pelo Decreto Estadual nº 7344 de 27.05.98
Comitê de Ética em Pesquisa – CEP / UESB

Jequié, 22 de dezembro de 2010

Of. CEP/UESB 440/2010

Ilma. Sra.
Profa. Dra. Maria Cátira Bortolini
Universidade Federal do Rio Grande do Sul - UFRGS

Prezado Senhor,

Comunicamos a V. Sª que o Projeto de Pesquisa abaixo especificado, foi analisado e considerado **APROVADO** pelo Comitê de Ética em Pesquisa – CEP/UESB, estando os pesquisadores liberados para o início da coleta de dados.

Protocolo nº: **212/2010**
CAAE: **0048.1.454.000-10**

Projeto: ***Fenótipo, ancestralidade genética e dinâmica de mestiçagem no Brasil***

Pesquisadores: ***Profa. Dra. Maria Cátira Bortolini (coordenadora)***
Caio Cesar Silva de Cerqueira , Profa. Tábita Hünemeier, Profa. Ana Angélica Leal Barbosa, Prof. Dr. Francisco Mauro Salzano (colaboradores)

Em atendimento à Resolução 196/96, deverá ser encaminhado ao CEP o relatório final da pesquisa (ver modelo no CEP), para acompanhamento pelo Comitê.

Atenciosamente,

Profª. Ivone Gonçalves Nery
Vice-Presidente do CEP/UESB

Av. José Moreira Sobrinho, S/N – Jequiezinho – Jequié-BA - Telefone: (73)3528-9727

6.4. ANEXO IV

(Questionário Individual aplicado aos voluntários do Projeto CANDELA)

PROJETO CANDELA

QUESTIONÁRIO INDIVIDUAL

Este questionário busca obter informações básicas sobre você, sobre suas impressões em relação a si mesmo e sobre suas opiniões em relação a certos aspectos desta pesquisa. Uma palavra utilizada com freqüência neste questionário é "ancestralidade". Utilizamos esta palavra em geral quando estamos falando sobre seus antepassados, os quais podem ser conhecidos ou desconhecidos. Em particular, neste questionário, ancestralidade significa a porcentagem de seus antepassados que eram de origem indígena, européia (branca) ou africana (negra). Se você não entender alguma outra palavra ou pergunta do questionário, por favor, sinta-se à vontade em perguntar aos pesquisadores responsáveis. Pedimos que, por favor, responda da maneira mais sincera possível. E não se preocupe que suas respostas são CONFIDENCIAIS.

Pesquisador Responsável:_____

Código: ZBR_____

Data: _____/_____/_____

1. **IDADE:_____**

2. **SEXO:** ( ) Feminino ( ) Masculino

3. **ESCOLARIDADE (por favor, marque qual seu último grau concluído):**

   ( ) PÓS-GRADUAÇÃO
   ( ) NÍVEL SUPERIOR
   ( ) FORMAÇÃO TÉCNICA - curso técnico
   ( ) ENSINO MÉDIO – até antigo 3º ano científico
   ( ) ENSINO FUNDAMENTAL - até antiga 8ª série
   ( ) NUNCA ESTUDOU

4. **Possui ou vive em casa própria?**
   ( ) Sim
   ( ) Não

5. **Renda mensal aproximada (em salários mínimos):_____**

6. **Para cada um dos seguintes itens ou serviços, marque a quantidade que você tem possui, se não possui marque zero (0):**

( ) 6.1. TV a cores
( ) 6.2. Rádio
( ) 6.3. Banheiro
( ) 6.4. Automóvel
( ) 6.5. Empregada Doméstica
( ) 6.6. Aspirador de pó
( ) 6.7. Lavadora de roupa
( ) 6.8. Videocassete/DVD
( ) 6.9. Refrigerador
( ) 6.10. Freezer
( ) 6.11. Computador
( ) 6.12. Lavadora de louça
( ) 6.13. Motocicleta
( ) 6.14. Bicicleta

7. **OCUPAÇÃO (por favor, marque no quadro abaixo dentre as 5 categorias a que mais se encaixa no seu trabalho atual ou, caso esteja desempregado, qual foi seu último trabalho):**

Prestador de Serviços	*Autônomo*	*Empregado*	*Servidor Público*	*Outro*
( ) Auxiliar de Serviços Gerais ( ) Técnico ( ) Responsável Técnico	( ) Trabalhador autônomo ( ) Agricultor ( ) Empresário	( ) Industrial ( ) Empregado com qualificação ( ) Gerenciamento	( ) Governo ( ) Professor ( ) Policial ou Militar	( ) Estudante ( ) Aposentado ( ) Estagiário

8. **Local de nascimento (cidade, estado e País):**

_____

9. **Local de nascimento (cidade, estado, país) dos seus antepassados:**

	Local	Língua (se diferente do português ou outra além do português)
Mãe		
Pai		
Avó materna		
Avô materno		
Avó paterna		
Avô paterno		

10. Alguma vez você se perguntou quem poderiam ser seus antepassados mais antigos?

( ) Sim
( ) Não

11. Você tem curiosidade sobre sua ancestralidade?

( ) Sim
( ) Não

12. Você sabe algo sobre estudos genéticos e ancestralidade?

( ) Sim
( ) Não

13. Você já ouviu falar em exames genéticos para conhecer sua ancestralidade?

( ) Sim
( ) Não

14. Se você sabe algo sobre estudos genéticos, como adquiriu esta informação? (marque todas as alternativas nas quais você adquiriu informações)

( ) Rádio, televisão
( ) Jornais, revistas, livros
( ) Palestras, atividades acadêmicas/escolares
( ) Internet
( ) Outra: _____

15. Você acha que os estudos de genética e ancestralidade podem ser úteis ou importantes?

( ) Sim
( ) Não
( ) Talvez
( ) Não sei

16. Você acha que os estudos de ancestralidade e genética podem ter riscos?

( ) Sim
( ) Não
( ) Talvez
( ) Não sei

17. Você já se perguntou sobre a origem dos seus antepassados?

( ) Sim
( ) Não

18. Você sabe se algum dos seus antepassados era:

	Sabe	Não sabe
Indígena?	( )	( )
Negro?	( )	( )
Europeu?	( )	( )

19. Se fosse possível, gostaria de saber sua ancestralidade?

( ) Sim
( ) Não
( ) Talvez
( ) Não sei

20. Você acha que existe racismo no Brasil?

( ) Sim
( ) Não

21. Você acha que este estudo pode ajudar a aumentar ou diminuir o racismo?

( ) Aumentar
( ) Diminuir
( ) Não sabe

22. Se você acha importante fazer um exame genético para conhecer a sua ancestralidade, por favor, marque EM CADA LINHA a alternativa que você está mais de acordo.

	Concordo totalmente	*Concordo parcialmente*	*Tanto faz*	*Discordo parcialmente*	*Discordo totalmente*
Para saber sobre seus antepassados	1	2	3	4	5
Para saber sobre sua saúde pessoal e familiar	1	2	3	4	5
Para saber sobre sua "nacionalidade"	1	2	3	4	5
Para ajudar a entender o passado	1	2	3	4	5
Para influenciar no futuro	1	2	3	4	5
Para saber sua identidade	1	2	3	4	5
Para saber mais sobre suas raízes	1	2	3	4	5
Por curiosidade	1	2	3	4	5
Outro (Por favor, descreva com suas palavras):					

23. Quais são os seus sentimentos em relação a sua ancestralidade? Por favor, marque EM CADA LINHA a alternativa que você está mais de acordo.

	Concordo totalmente	*Concordo parcialmente*	*Tanto faz*	*Discordo parcialmente*	*Discordo totalmente*
Orgulho	1	2	3	4	5
Felicidade	1	2	3	4	5
Vergonha	1	2	3	4	5
Indiferença	1	2	3	4	5
Tristeza	1	2	3	4	5
Curiosidade	1	2	3	4	5
Outro (Por favor, descreva com suas palavras):					

24. Você se considera:

( ) Negro
( ) Mulato
( ) Indígena
( ) Moreno
( ) Mestiço
( ) Branco
( ) Pardo
( ) Europeu
( ) Outro:_____

25. Você acha que existe alguma relação entre sua aparência e ancestralidade?

( ) Sim
( ) Não
( ) Talvez
( ) Não sei

26. O que você acha que afeta sua aparência física? (Em cada linha, marque EM CADA LINHA a alternativa que você está mais de acordo)

	Nada	*Afeta pouco*	*Afeta bastante*
Lugar de nascimento	0	1	2
A "raça" dos seus antepassados	0	1	2
Seu estilo de vida	0	1	2
Sua alimentação	0	1	2
Seu trabalho	0	1	2
Sua saúde	0	1	2
Quanto dinheiro você possui	0	1	2
Seu comportamento	0	1	2
Suas crenças	0	1	2

27. Qual você acha que é sua ancestralidade? Ou seja, a porcentagem aproximada dos seus antepassados que foram de origem indígena, negra/africana ou européia.

Marque com um X a categoria de porcentagem que você considera ter para cada ancestralidade. Se achar que não possui uma determinada ancestralidade, **marque a categoria 0-20%.**

Indígena		**Negra/Africana**		**Européia**	
0 – 20% (nula ou muito baixa)	( )	0 – 20% (nula ou muito baixa)	( )	0 – 20% (nula ou muito baixa)	( )
20 – 40% (baixa)	( )	20 – 40% (baixa)	( )	20 – 40% (baixa)	( )
40 – 60% (média)	( )	40 – 60% (média)	( )	40 – 60% (média)	( )
60 – 80% (alta)	( )	60 – 80% (alta)	( )	60 – 80% (alta)	( )
80 – 100% (muito alta ou total)	( )	80 – 100% (muito alta ou total)	( )	80 – 100% (muito alta ou total)	( )

Se você acha que possui outros tipos de ancestralidade, por favor, nos diga quais:

_____

MUITO OBRIGADO PELA PARTICIPAÇÃO!

6.5. ANEXO V

(Ficha de informações confidenciais aplicada aos voluntários do Projeto CANDELA)

CANDELA

Consórcio para Análise da Diversidade e Evolução da América Latina

FICHA DE INFORMAÇÃO CONFIDENCIAL

DATA: _____/_____/_____

CÓDIGO DO VOLUNTÁRIO: ZBR_____

Nome completo:_____

Data de nascimento (dd/mm/aaaa):_____/_____/_____

Local de nascimento:_____
(cidade, estado e país)

Endereço:_____

_____

Telefone: (____)_____

Celular: (____)_____

E-mail:_____

6.6. ANEXO VI

(Questionário complementar aplicado aos voluntários do Projeto CANDELA)

QUESTIONÁRIO COMPLEMENTAR

ZBR:_____

1A. NOME COMPLETO e NACIONALIDADE dos parentes (caso saibam):

Nome do Pai:_____

NACIONALIDADE:_____

Nome da Mãe:_____

NACIONALIDADE:_____

Nome do Avô Paterno:_____

NACIONALIDADE:_____

Nome da Avó Paterna:_____

NACIONALIDADE:_____

Nome do Avô Materno:_____

NACIONALIDADE:_____

Nome da Avó Materna:_____

NACIONALIDADE:_____

2A. Você tem conhecimento sobre a ocorrência de doenças genéticas em seus parentes? Em caso afirmativo, qual doença seria? Há registro de algum caso de câncer em parentes de primeiro e segundo graus? Qual tipo?

6.7. ANEXO VII

(Termo de consentimento livre e esclarecido assinado pelos voluntários do Projeto CANDELA na UFRGS/HCPA)

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Resolução nº 196, de 10 de Outubro de 1996, sendo o Conselho Nacional de Saúde.

O presente termo em atendimento à Resolução 196/96, destina-se a esclarecer ao participantes aspectos relevantes da pesquisa intitulada "**Fenótipo, ancestralidade genômica e dinâmica de miscigenação no Brasil**" sob responsabilidade dos pesquisadores **Caio Cesar Silva de Cerqueira, Tábita Hünemeier e Dânae Longo**, do Programa de Pós-Graduação em Genética e Biologia Molecular em nível de Doutorado e Pós-Doutorado, respectivamente, bem como **Lavínia Schüler-Faccini** e **Maria Cátira Bortolini**, docentes do Serviço de Genética Médica (LSF) e do Departamento de Genética (LSF e MCB), ambas da Universidade Federal do Rio Grande do Sul .

Gostaríamos de convidá-lo a participar do estudo "Fenótipo, ancestralidade genômica e dinâmica de mestiçagem no Brasil" que busca correlacionar características fenotípicas, tais como cor da pele, olhos, cabelo e morfologia do crânio com dados genéticos, bem como com informações pessoais e de terceiros sobre pertencimento a grupos sociais. Temos como meta avaliar os vários fatores que podem influenciar a diversidade fenotípica normal dentro de nossa espécie e como essa é percebida em nossa sociedade. Para isso será coletado material biológico (20ml de sangue através de punção venosa com seringas descartáveis) para posterior extração de DNA (ácido desoxirribonucléico). O DNA será obtido com a utilização de um Kit comercial QIAamp DNA Mini Kit (pequeno tubo). Agulhas descartáveis serão utilizadas e a coleta será realizada por técnico capacitado em ambiente apropriado. Tal procedimento pode causar algum desconforto no local da picada. O DNA será extraído no Laboratório de Evolução Molecular e Genética de Populações Humanas do Departamento de Genética da UFRGS, onde também ficará estocado por cinco anos conforme previsto pela RESOLUÇÃO No 347, DE 13 DE JANEIRO DE 2005 do CONSELHO NACIONAL DE SAÚDE.

O material será analisado através de métodos laboratoriais específicos para caracterizar os genes relacionados com as características morfológicas de interesse. Dados adicionais serão coletados a partir de:

(1) Fotografias faciais, medidas antropométricas (a saber: peso, altura, índice cefálico, largura mandibular), medida constitutiva da pigmentação da pele (aferida através de um aparelho chamado espectrofotômetro), dos cabelos e olhos. As fotos serão utilizadas para obtenção de parâmetros craniofaciais. Também poderemos utilizar sua foto anonimamente para que outros voluntários manifestem impressões que dizem respeito à sua ancestralidade, grupo étnico ou grupo de cor. As fotos não serão disponibilizadas em *web sites*, nem em publicações científicas e outras de qualquer natureza.
(2) Questionários e entrevistas estruturadas para documentar o status sócio-econômico, auto-percepção da ancestralidade e a percepção do investigador, assim como informação

sobre a visão dos participantes sobre a natureza da "raça" e sobre a pesquisa propriamente dita.

Finalmente a análise de dados incluirá: (I) estimativa de ancestralidade individual; (II) avaliação da associação de ancestralidade genética com as medidas fenótipicas e com as variáveis sociológicas coletadas; (III) avaliação do efeito da pesquisa (incluindo a estimativa da ancestralidade genética), sobre os conceitos pré-existentes dos indivíduos.

Todo material coletado será armazenado no Laboratório de Genética e Biologia Molecular do Departamento de Genética da Universidade Federal do Rio Grande do Sul, sem descrição de nome, e com um número de código com chave de conhecimento exclusivo dos pesquisadores. Caso haja interesse futuro na utilização dos dados obtidos para outras linhas de investigação, os pacientes somente serão incluídos após a assinatura de um novo termo de consentimento.

Pelo presente consentimento informado declaro que fui esclarecido, de forma clara e detalhada:

- da garantia de receber resposta a qualquer pergunta ou esclarecimento a qualquer dúvida a cerca dos procedimentos, riscos ou benefícios e outros assuntos relacionados com a pesquisa;
- da segurança de que não serei identificado e que se manterá o caráter confidencial das informações relacionadas com a minha privacidade;
- da liberdade de retirar meu consentimento informado, a qualquer momento, e deixar de participar do estudo.

- **Consentimento para participação:** Eu estou de acordo com a participação no estudo descrito acima. Eu fui devidamente esclarecido quanto os objetivos da pesquisa, aos procedimentos aos quais serei submetido e os possíveis riscos envolvidos na minha participação. Os pesquisadores me garantiram disponibilizar qualquer esclarecimento adicional que eu venha solicitar durante o curso da pesquisa e o direito de desistir da participação em qualquer momento, sem que a minha desistência implique em qualquer prejuízo à minha pessoa ou à minha família, sendo garantido anonimato e o sigilo dos dados referentes a minha identificação, bem como de que a minha participação neste estudo não me trará nenhum benefício econômico.

Eu, _____, aceito livremente participar do estudo intitulado "Fenótipo, ancestralidade genômica e dinâmica de miscigenação no Brasil" **desenvolvido pelos pesquisadores Caio Cesar Silva de Cerqueira, Tábita Hünemeier e Danae Longo, sob a responsabilidade das Professoras Maria Cátira Bortolini e Lavínia Schuler Faccini da Universidade Federal do Rio Grande do Sul (UFRGS).**

Assinatura do Participante_____ Data: ____/_____/____

Assinatura da pessoa ou responsável legal_____ Data: ____/_____/____

COMPROMISSO DO PESQUISADOR

Eu discuti as questões acima apresentadas com cada participante do estudo. É minha opinião que cada indivíduo entenda os riscos, benefícios e obrigações relacionadas a esta pesquisa.

_____,
Assinatura do Pesquisador

_____, Data: ____/_____/____
Nome do Pesquisador

Para maiores informações, pode entrar em contato com:

Maria Cátira Bortolini. Departamento de Genética, Universidade Federal do Rio Grande do Sul, Campus do Vale, Av. Bento Gonçalves, no. 9500, prédio 43323, sala 126. Fone: (51) 3308 9844
Lavinia Schuler Faccini. Fone (51) 3308-9826
Caio Cesar Silva de Cerqueira. Fone: (51) 3308 6746
Tábita Hünemeier. Fone: (51) 3308 6746
Dânae Longo. Fone: (51) 3308-9819
Comitê de Ética da Universidade Federal do Rio Grande do Sul. Fone: (51) 3308 3738.
Comitê de Ética do Hospital de Clínicas de Porto Alegre: Telefone para contato: 3359 7640.

Comitê de Ética em Pesquisa
GPPG/HCPA
VERSÃO APROVADA
28/11/2011
nº 100565

6.8. ANEXO VIII

(Termo de consentimento livre e esclarecido assinado pelos voluntários do Projeto CANDELA na UESB/ Jequié-BA)

UNIVERSIDADE ESTADUAL DO SUDOESTE DA BAHIA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Resolução nº 196, de 10 de Outubro de 1996, sendo o Conselho Nacional de Saúde.

O presente termo em atendimento à Resolução 196/96, destina-se a esclarecer ao participante da pesquisa intitulada "**Fenótipo, ancestralidade genômica e dinâmica de mestiçagem no Brasil"** sob responsabilidade dos pesquisadores "**Caio Cesar Silva de Cerqueira"**, do curso de Pós-Graduação em Genética e Biologia Molecular em nível de Doutorado e "**Maria Cátira Bortolini"**, Docente do Departamento de Genética, ambos da Universidade Federal do Rio Grande do Sul, os seguintes aspectos:

A presente pesquisa que tem como título "Fenótipo, ancestralidade genômica e dinâmica de mestiçagem no Brasil" busca correlacionar características físicas, tais como cor da pele, olhos, cabelo e morfologia do crânio com dados genéticos, bem como com informações pessoais e de terceiros sobre pertencimento a grupos sociais. Temos como meta avaliar os vários fatores que podem influenciar a diversidade física normal dentro de nossa espécie e como essa é percebida em nossa sociedade. Para isso será coletado material biológico (saliva) para posterior extração de DNA. A coleta de saliva será em pequena quantidade (600 microlitros) após escovação. Para voluntários que se dispõe também será coletado 20ml de sangue através de punção venosa. O DNA será obtido com a utilização de um Kit comercial QIAamp DNA Mini Kit (pequeno tubo). Agulhas descartáveis serão utilizadas e a coleta será realizada por técnico capacitado em ambiente apropriado. Tal procedimento pode causar algum desconforto no local da picada. O DNA será extraído no Laboratório de Evolução Molecular e Genética de Populações Humanas do Departamento de Genética da UFRGS, onde também ficará estocado por cinco anos conforme previsto pela RESOLUÇÃO No 347, DE 13 DE JANEIRO DE 2005 do CONSELHO NACIONAL DE SAÚDE.

O material será analisado através de métodos laboratoriais específicos para caracterizar os genes relacionados com as características morfológicas de interesse. Dados adicionais serão coletados a partir de:

(1) Fotografias faciais, medidas antropométricas (i.e. peso, altura, índice cefálico, largura mandibular), medida constitutiva da pigmentação da pele (aferida através de um aparelho chamado espectrofotômetro), dos cabelos e olhos. As fotos serão utilizadas somente para obtenção de parâmetros craniofaciais e não serão disponibilizadas em web sites, nem em publicações científicas e outras de qualquer natureza.
(2) Questionários e entrevistas estruturadas para documentar o status sócio-econômico, auto-percepção sobre a região de origem dos ancestrais e a percepção do investigador, assim como informação sobre a visão dos participantes sobre a natureza da "raça" e sobre a pesquisa propriamente dita.

Finalmente a análise de dados incluirá: (i) estimativa com dados do DNA dos componentes africano, europeu e indígena de cada indivíduo; (ii) A estimativa genética do componente africano, europeu e indígena de cada indivíduo será comparado com as medidas fenótipicas e com as variáveis sociológicas coletadas; (iii) avaliação do efeito da pesquisa (incluindo a estimativa dos componentes africanos, europeu e indígena), sobre os conceitos pré-existentes dos indivíduos.

Todo material coletado será armazenado no Laboratório de Genética e Biologia Molecular do Departamento de Genética da Universidade Federal do Rio Grande do Sul, sem descrição de nome, e com um número de código com chave de conhecimento exclusivo dos pesquisadores. Caso

haja interesse futuro na utilização dos dados obtidos para outras linhas de investigação, os pacientes somente serão incluídos após a assinatura de um novo termo de consentimento.

Pelo presente consentimento informado declaro que fui esclarecido, de forma clara e detalhada:

- da garantia de receber resposta a qualquer pergunta ou esclarecimento a qualquer dúvida a cerca dos procedimentos, riscos ou benefícios e outros assuntos relacionados com a pesquisa;
- da segurança de que não serei identificado e que se manterá o caráter confidencial das informações relacionadas com a minha privacidade;
- da liberdade de retirar meu consentimento informado, a qualquer momento, e deixar de participar do estudo.

- **Consentimento para participação:** Eu estou de acordo com a participação no estudo descrito acima. Eu fui devidamente esclarecido quanto os objetivos da pesquisa, aos procedimentos aos quais serei submetido e os possíveis riscos envolvidos na minha participação. Os pesquisadores me garantiram disponibilizar qualquer esclarecimento adicional que eu venha solicitar durante o curso da pesquisa e o direito de desistir da participação em qualquer momento, sem que a minha desistência implique em qualquer prejuízo à minha pessoa ou à minha família, sendo garantido anonimato e o sigilo dos dados referentes a minha identificação, bem como de que a minha participação neste estudo não me trará nenhum benefício econômico.

Eu, _____, aceito livremente participar do estudo intitulado "Fenótipo, ancestralidade genômica e dinâmica de mestiçagem no Brasil**" desenvolvido pelos(as) acadêmicos(as) Caio Cesar Silva de Cerqueira e Tábita Hünemeier, sob a responsabilidade do(a) Professor(a) Maria Cátira Bortolini da Universidade Federal do Rio Grande do Sul (UFRGS).**

Nome do Participante_____

Nome da pessoa ou responsável legal_____

Polegar direito

COMPROMISSO DO PESQUISADOR

Eu discuti as questões acima apresentadas com cada participante do estudo. É minha opinião que cada indivíduo entenda os riscos, benefícios e obrigações relacionadas a esta pesquisa.

_____ Jequié, Data: ____/_____/____

Assinatura do Pesquisador

Para maiores informações, pode entrar em contato com:

Maria Cátira Bortolini. Fone: (51) 3308 9844
Caio Cesar Silva de Cerqueira. Fone: (51) 3308 6746
Tábita Hünemeier. Fone: (51) 3308 6746
Comitê de Ética da Universidade Federal do Rio Grande do Sul. Fone: (51) 3308 3738.

6.9. ANEXO IX

(Ficha fenotípica aplicada aos voluntários do Projeto CANDELA)

CANDELA
PHENOTYPIC DATA SHEET

ID: _____

	1st	2nd	3rd
Weight *(kg)*			
Stature (height) *(cm)*			
Hip circumference *(cm)*			
Waist circumference *(cm)*			
Chelion-chelion breadth *(mm)*			
Nasion-gnation height *(mm)*			
Head size *(mm)*			

Melanin index (skin reflectance) / Inner surface of the upper arm

Right: _____ Left: _____

Eye colour

blue/grey ( )
green ( )
honey ( )
light brown ( )
dark brown/black ( )

Balding

no hair loss ( )
frontal baldness ( )
(+) mild vertex baldness ( )
(+) moderate vertex baldness ( )
(+) severe vertex baldness ( )

Natural hair colour

red/reddish ( )
blond ( )
dark blond/light brown ( )
brown/black ( )

Greying

no greying ( )
predominant no-greying ( )
~50% greying ( )
predominant greying ( )
totally withe hair ( )

Hair morphology

straight ( )
wavy ( )
curly ( )
afro ( )

Handedness

right ( )
left ( )

Craniofacial malformation	(Y) (N)	_____
Craniofacial trauma	(Y) (N)	_____
Hormonal treatment	(Y) (N)	_____
Surgery or orthodontic	(Y) (N)	_____
Other		_____

Additional information:

Name of the anthropometrist:
Date:

7. APÊNDICES

7.1. APÊNDICE I

(Metodologia de coleta de amostra empregada nesta tese)

O recrutamento dos voluntários para o desenvolvimento do Projeto CANDELA-Brasil se iniciou em outubro do ano de 2010 e findou-se em dezembro de 2012. Neste período também foram realizadas as coletas nos outros quatro países participantes (Peru, Chile, Colômbia e México). Cada país teve o objetivo de coletar 1.500 voluntários para a pesquisa, totalizando 7.500. As 563 amostras utilizadas para o desenvolvimento desta tese fazem parte do projeto supramencionado.

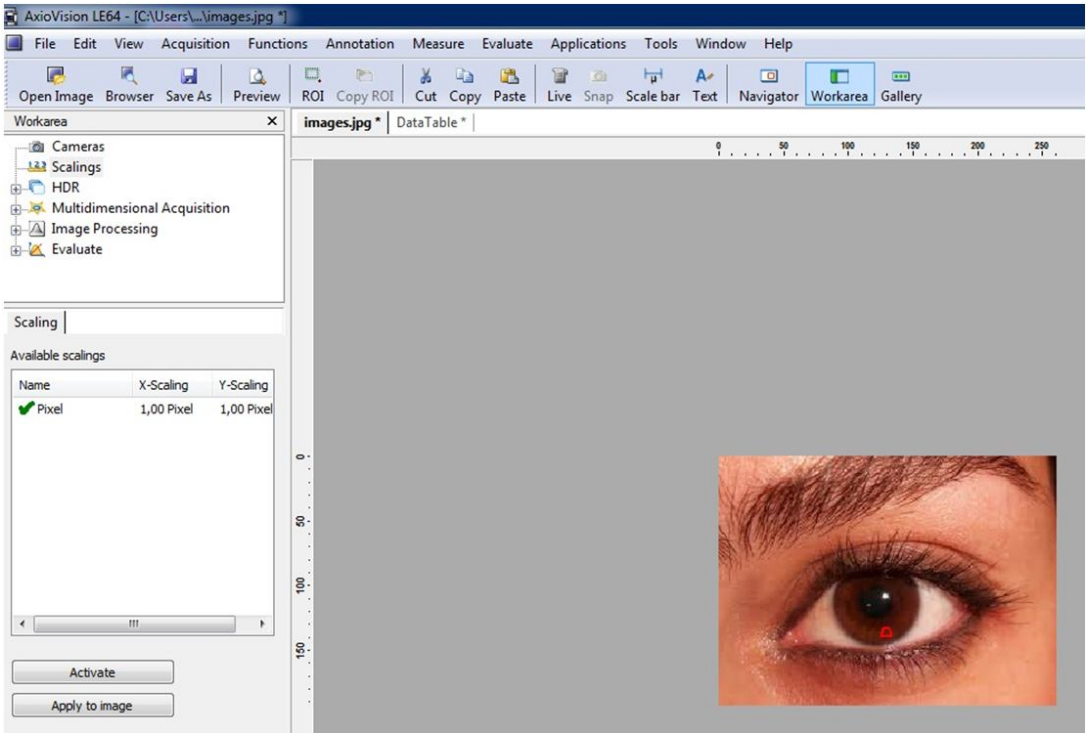
O protocolo de coletas é o mesmo para cada país participante e envolveu tomada de medidas antropométricas (peso; altura; índice de melanina; circunferências do quadril, cintura e cabeça; largura e comprimento da cabeça; largura da boca; distância do nasion-gnation; medida do índice de melanina; qualificação da cor dos olhos e cabelos; morfologia dos cabelos; quantidade aproximada de grisalho; quantidade aproximada de calvície; bem como outras informações indicadas no Anexo IX), tomada de fotos e preenchimento de outros questionários do Projeto CANDELA (Anexos IV a VIII). As fotografias foram tomadas da cabeça dos voluntários em 5 ângulos por três vezes: lateral direita (90°), lateral esquerda (90°), frontal (0°), fronto-lateral direita (45°) e fronto-lateral esquerda (45°). No Brasil, os pesquisadores integrantes do consórcio CANDELA envolvidos nas coletas foram a Dra. Tábita Hünemeier, Dra. Virgínia Ramallo, e o Doutorando Caio Cesar Silva de Cerqueira.

Com relação às características de pigmentação, conforme especificado no Anexo IX, a cor dos olhos e cabelos dos voluntários foram qualificadas em 5 e 4 categorias, respectivamente. Para a cor dos olhos, as categorias foram: Azul/Cinza, Verde, Mel, Castanho Claro, Castanho Escuro/Preto. Para a cor dos cabelos, as categorias foram: Ruivo, Loiro, Loiro Escuro/Castanho Claro, Castanho Escuro/Preto. A cor da pele dos voluntários, por sua vez, foi quantificada com o aparelho mostrado na Figura 4 (*DSMII* ColorMeter) do corpo desta tese, utilizando o sistema E & M (Eritema e Melanina). O Índice de Melanina (IM) foi medido abaixo dos braços direito e esquerdo. Este índice varia de 20 (cor de pele mais clara) a 100 (cor de pele mais escura). Para as análises, foi utilizada uma média aritmética destes valores para cada voluntário. Mais detalhes deste procedimento, consultar a ‘Introdução’ desta tese.

Exclusivamente no Brasil, foram feitas também as quantificações da cor dos olhos e cabelos das amostras, especificamente para os voluntários analisados nesta tese. Esta quantificação utilizou o *software* AxioVision LE 4.8.2 (Disponível em: http://applications.zeiss.com//C1257A26004B6E67/allBySubject/C57C9DEA76A98584C1257A2600643CFC). Para esta quantificação, foram utilizadas as fotografias dos voluntários,

tomadas no momento das coletas. No *software* AxioVision LE 4.8.2 era aberta a fotografia do voluntário (preferencialmente a fotografia frontal para medida da cor dos olhos; ou lateral esquerda ou direita para a cor dos cabelos), aproximando-se o máximo possível para a medida (conforme Figura 7.1a). Em seguida, era realizada um marco na figura (marcado em vermelho na Figura 7.1a) e era pedido ao *software* os valores de RGB (Red-Green-Blue) da região delimitada pelo círculo vermelho (Figura 7.1b). Para a medida da cor dos olhos, foram utilizadas regiões mais periféricas da íris, a fim de evitar as manchas oculares observadas em alguns indivíduos. Com relação à cor dos cabelos, utilizou-se regiões mais próximas ao couro cabeludo, a fim de evitar regiões dos cabelos mais desgastadas pelo sol ou regiões possivelmente tingidas. No caso da quantificação dos cabelos, alguns voluntários foram excluídos para medida, pois possuíam cabelos bastante grisalhos ou traziam bastantes dúvidas quanto ao tingimento total dos cabelos. Este procedimento de quantificação foi feito para o olho esquerdo e direito de cada um dos 563 voluntários do CANDELA-Brasil, bem como para os cabelos. Mais informações sobre estes procedimentos, favor consultar texto.

a)



b)

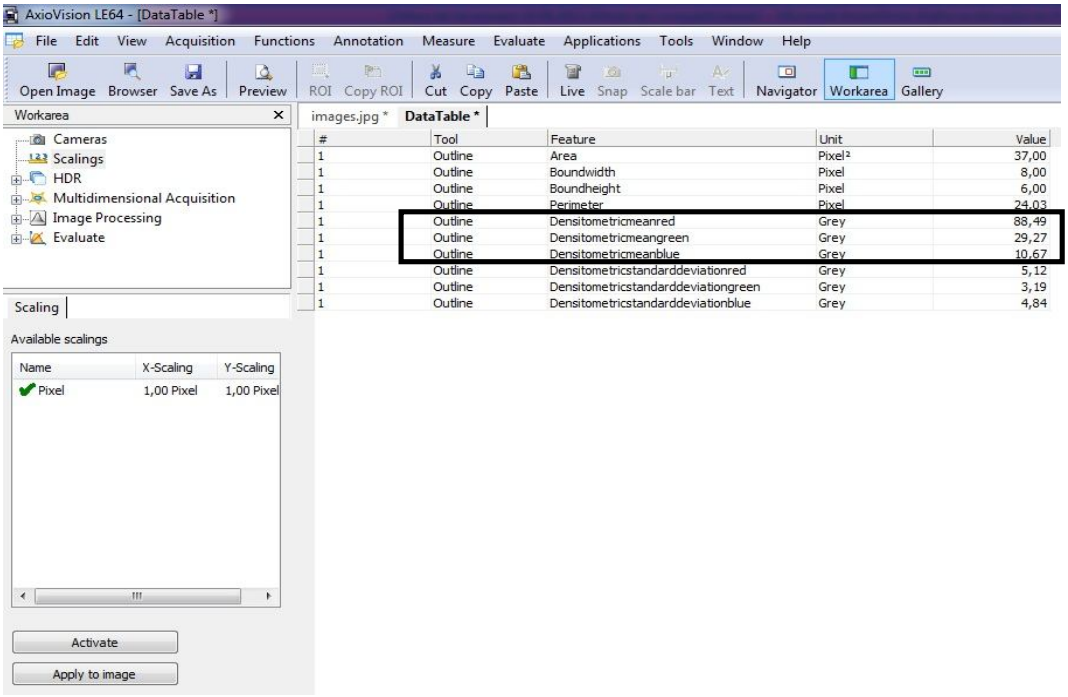


Figura do Apêndice 7.1. Etapas na medida quantitativa da cor dos olhos. a) imagem dos olhos de uma foto extraída do google como exemplo. Note o pequeno círculo vermelho delimitando a região medida. b) *Output* dos valores de RGB mensurados a partir do círculo vermelho. Os valores de *Red*, *Green* e *Blue* estão salientados dentro do retângulo. Para a cor dos cabelos, etapas similares foram seguidas. Para mais informações favor consultar texto.

7.2. APÊNDICE II

(Capítulo “Human pigmentation genes: forensic perspectives, general aspects and evolution”, do livro “Forensic Science”, publicado pela Editora *Nova Science Publishers*, 2011)

In: Forensic Science
Editors: N. Yacine and R. Fellag, pp.
ISBN 978-1-61324-999-4

Chapter 3

HUMAN PIGMENTATION GENES: FORENSIC PERSPECTIVES, GENERAL ASPECTS AND EVOLUTION

Caio Cesar Silva de Cerqueira[a], Carlos Eduardo Guerra Amorim[a], Francisco Mauro Salzano[a], Maria Cátira Bortolini[a,*]

[a]Genetics Department, Biosciences Institute, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

ABSTRACT

Despite the growing number of registered DNA forensic profiles, the rate of profile hits is far below the expected. When a given profile is not found with the standard kits, any available information is important. Ancestry Informative Markers (AIM) have been employed for the identification of the person's ethnicity. However, this is not an adequate approach in certain cases, as, for instance, individuals living in ethnically admixed populations, in which the physical appearance is not necessarily associated with ethnicity. Hence, the use of AIMs in such populations has few or no advantage. The interest in the search for alleles directly linked to physical characteristics is therefore rapidly increasing, resulting in the commercialization of kits for the identification of skin, hair and eye pigmentation phenotypes based on genetic variation of candidate genes such as *ASIP, HERC2, MC1R, OCA2,* and *TYR*. This genetic variation may affect different stages of the pigmentation process, including melanogenesis, the stabilization and transport of enzymes during melanin synthesis, melanosome production and maintenance, and the balance between the synthesis of different types of melanin. In this chapter we examine the most important genes associated to human hair, skin and eye pigmentation, discussing the evolutionary background for the observed variation and their functional relevance for the physiological mechanisms involved with pigmentation, highlighting the forensic application of this knowledge. We will also discuss better ways of collecting pigmentation phenotypic data in human populations for forensic and general studies. It is expected that in the near future we may be able to predict with high reliability the externally visible human characteristics based on DNA analyses.

* Correspondence: Dr. Maria Cátira Bortolini, Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. Email: maria.bortolini@ufrgs.br; Telephone: +55 (51) 3308-9844.

1. INTRODUCTION

1.1. General Considerations

With the advent of forensic genetics and attempts to include information of phenotypic characteristics of possible suspects in police databases, it is desirable to generate data about these traits and additionally to explore and integrate the information generated by the scientific community (see recent discussion in Enserink, 2011). However, it is necessary to uncover the relationship of each gene in particular with its phenotype, something nontrivial due to the enormous complexity of the genotype → phenotype connections. Specific morphological characteristics of plants, animals or humans are products of processes that involve the spatial and temporal expression of several genes that interact during development, despite the existence of variants in genes of major effect, which may lead to great morphological changes (Major gene effect hypothesis; Nei, 2007). The indicated major gene effect hypothesis suggests that morphological evolution can occur by the action of a small number of mutations of large effect on structural and/ or regulatory genes. Thus, when dealing with human pigmentation, it is expected that several genes may be involved in this trait, as well as environmental and epigenetic factors.

Many hypotheses predict how we can find pigmentation variation over human populations. For many authors, skin color is one of the most illustrative examples of the action of natural selection. This factor would be responsible for the elimination of genetic variants associated with light skin in areas with high levels of ultraviolet radiation, as a form of protection to sun damage (i.e. burns, malignant melanoma, basal and squamous cell carcinoma) (Blum, 1961; Kollias et al., 1991). The photoprotective properties of a highly melanized skin and the recent African origin of modern humans suggest that the ancestral phenotype would be dark skin (Jablonski & Chaplin, 2000; Rogers et al., 2004). Thus, the appearance of white skin would have occurred after the departure of modern humans from Africa, and two possibilities exist for its increase in frequency: (a) relaxation of functional constraints, and drift of the derived alleles (Brace, 1963); or (b) in regions of low UV radiation, positive selection favoring the fixation of mutations that resulted in a clear skin as a way to maximize cutaneous vitamin D synthesis (Rana et al., 1999; Jablonski & Chaplin, 2000).

2. PIGMENTATION AND GENETIC VARIATION

2.1. Mechanisms of Development

Human hair, eye, and skin pigmentation are the most visible examples of the human phenotypic diversity (Sulem et al., 2007). Since Darwin's first studies, it was believed that human pigmentation variation could be a consequence of natural selection (Darwin, 1871). Skin color is associated to latitude, with people having increasingly fairer skin as far as they live away from the equator (Relethford, 1997). On the other hand, the most diverse eye and hair pigmentation is found in the European and Euro-derived populations. The others present mainly brown eyes and dark hair. In contrast to skin color, the adaptive significance of a loss

of eye pigmentation is less evident. The high degree of exposure to light of the white region of the eyeball around the iris is unique to humans and has been proposed as a factor to enhance signals stare, for example (Kobayashi & Kohshima, 2001).

One suggestion that seems to have little support is that the iris pigmentation could influence visual acuity in low-light environments. Although albinism is an exception, with obvious eye problems due to loss in the ability of melanin synthesis, there are reports of the effects of eye color on reaction time and involvement in specific sports (Rowe & Evans, 1994), associations with shyness in childhood (Coplan et al., 1998), links with hearing loss (Da Costa et al., 2008), and sexual selection (Frost, 2006). The latter may have been a cofactor for the loss of pigmentation in humans. Haplotype analyses suggest that the blue eye color has originated about 10,000 years ago, independently of albinism, as a founder mutation shared by several European populations (Eiberg et al., 2008). There are also reports that the ability to recover from the seasonal affective disorder (SAD), a major depressive illness, is linked to blue eyes (Goel et al., 2002; Terman & Terman, 1999). Extrapolating these results, perhaps individuals with blue eyes might have been able to better adapt to darkness in Neolithic Europe gloomy winter's day better than those with brown eyes.

The organic pigment melanin is synthesized and stored inside melanosomes in the melanocytes – dendritic cells located between dermis and epidermis in skin, in the hair bulb and in the iris (Hida et al., 2009). During ontogenesis primary melanocytes – the melanoblats – migrate from the neural crest to the skin, to hair follicles and to eyes, and are directly responsible for the pigmentation of such regions (Sturm et al., 2001). The variation in human pigmentation is associated to melanin type, number, size, and density of melanosomes, and also their pH. There are two main types of melanin: feomelanin (red – yellow) and eumelanin (black – brown), which are synthesized according to the chemical and hormonal environment present inside the melanosomes (Barsh, 2006). Tanning, for instance, is associated to eumelanin production, which is synthesized in response to ultraviolet radiation (Nan et al., 2009a), to protect the skin from damage. The melanosome's pH works as a regulatory factor for melanin synthesis (Cheli et al., 2009; Lamason et al., 2005).

The number of melanocytes does not appear to differ between different eye colors (Imesch et al., 1997). However, it has been reported that the total number of melanocytes in the iris may be lower in Asians as compared to Africans and Europeans, due to the smaller area of the iris or low density of melanocytes in that group (Albert et al., 2003). Unlike skin and hair, where melanin is continuously produced and secreted, in the iris melanosomes are retained and accumulated in the melanocytes cytoplasm within the iris stroma. Melanin type has also been studied chemically between eye colors, and blue iris is associated with minimal pigment content, while the eumelanic and pheomelanic forms were detected in other eye colors (Prota et al., 1998; Wielgus & Sarna, 2005). Thus, change in packaging, the quality and quantity of melanin pigment is what causes the observed spectrum of the eyes, hair and skin hues.

The pigmentation candidate genes that were identified up to the present act during many different stages of the process, including melanogenesis, stabilization and transport of enzymes during melanin synthesis, melanosome production and maintenance, and balance between the synthesis of different melanin types (Norton et al., 2007). The basic mechanisms of human pigmentation regulation involve a series of processes. Melanocites respond to the peptide alfa-MSH hormones (alfa-melanocite stimulant hormone) and to ACTH (adenocorticotropic hormone) through a protenin-G-coupled receptor (Mc1r) to drive melanin

production, which is derived from the enzymatic oxidation of tyrosine (Sturm et al., 2001). Alfa-MSH binds to the product of the *MC1R* gene, increasing the intracellular cyclic adenosine monophosphate (cAMP, Buscà & Ballotti, 2000) and activating the microphthalmia transcription factor (Mitf; Goding, 2000). It activates many enzymes in the melanosomes, the first being tyrosinase (Tyr; Marks & Seabra, 2001), which acts on tyrosine to produce dopaquinone; the latter links to cysteine molecules, if they are available in the cellular environment. After these interactions, the product of this reaction (cysteinylDOPA) is then oxidated to produce feomelanin (Ito, 2003). The *MITF* gene product also stimulates the tyrosine-related protein (*Tyrp1*) and the dopachrome tautomerasis (*Dct*; Marks & Seabra, 2001). Without cysteine, the dopainones found at the melanocites are transformed in other compound to produce eumelanin through the action of *Tyrp1* and *Dct* (Ito, 2003). The *Pmel17, Matp, Oca2*, and *Nckx5* proteins are all required for optimal melanin synthesis. *Pmel17* forms the fibrilar matrix where eumelanin is formed (reviewed by Theos et al., 2005). The *SLC45A2, OCA2,* and *SLC24A5* gene products are involved in inter-membrane transportation (Lee et al., 1995; Newton et al., 2001, Lamason et al., 2005), driving eumelanin production and eumelanosome maturation. The *ASIP* gene product binds to the *MC1R* gene product, blocking its association to alfa-MSH and modifying melanin production. Many genome-wide association studies (GWAS) indicated the importance of the above-mentioned genes through the identification of single-nucleotide polymorphisms (SNPs) and other types of genetic variation (Pho & Leachman, 2010).

A study of *SLC24A5*, *SLC45A2* and *TYR* polymorphisms found highly significant associations between their variation and skin reflectance measurements, an indirect measure of melanin content (Stokowski et al., 2007). The additive association found in these three genes accounts for a large fraction of the natural variation in skin pigmentation found in South Asian populations. This was the first study analyzing polymorphisms across the genome to find possible genetic determinants of natural variation in pigmentation within a human population. Norton et al. (2007), on the other hand, showed evidence that at least two other genes (*ASIP* and *OCA2*) work in shaping light and dark pigmentation around the globe, and also identified several other candidate loci that have a significant effect on regional pigmentation phenotypes worldwide. The data of Norton et al. (2007) are important because they showed independent genetic effects for light skin in Europeans and East Asians, which probably arose after their divergence. Details of the aforementioned genes, and other possibly involved in pigmentation, as well as some of their variants, will be discussed in greater detail in the following items.

2.1. *MC1R*

One of the most important systems regulating human pigmentation is the melanocortin 1 receptor (*MC1R*), G-protein coupled (Sturm, 2009). The genetic mechanisms of normal variation in skin color were only better understood after the discovery of mutations in this gene, encoding the Mc1r receptor (Makova & Norton, 2005, Valverde et al., 1995). The *MC1R* gene is located on chromosome 16q24.3 (Gantz et al., 1994), has 5 exons in humans, and its product is a cell surface receptor of melanocytes, working towards the production of melanin. It also participates in the regulation of eumelanin and pheomelanin production. The gene produces a 317 aminoacids protein (Sturm, 2009) and is polymorphic (Box et al., 1997;

Rana et al., 1999). Many of its variants are related to red hair and fair skin, increased risk for skin cancer and poor tanning ability, and it shows interactions with the *OCA2* gene (King et al., 2003).

Nine common *MC1R* alleles have been analyzed through the expression of the variant protein in studies in vitro (Beaumont et al., 2007). The Arg151Cys and Arg160Trp polymorphisms (rs1805007, rs1805008) produce changes in the cellular localization of the receptor encoded by the *MC1R* gene (Schioth et al., 1999; Beaumont et al., 2005). Another polymorphism, Asp294His (rs1805009), interferes with the coupling ability of the G protein in the cell membrane surface (Schioth et al., 1999; Duffy et al., 2004). In a GWAS study conducted by Nan et al. (2009a), a highly significant association between some SNPs outside the *MC1R* gene on chromosome 16 and the hair color phenotype, disappear after adjustment for functional variants (Arg151Cys, Arg160Trp, Asp294His) of the *MC1R* gene.

2.2. *OCA2*

The *OCA2* gene has 24 exons and is located on chromosome 15q11.2-15q12 (Sturm & Frudakis, 2004), produces a protein with 838 aminoacids (Sturm, 2009) and its function is not well established (Rebbeck et al., 2002): it was suggested to be a channel (antiporter) of Na^+/H^+ (Ancans et al., 2001; Puri et al., 2000) or a glutamate transporter (Lamoreux et al., 1995). Both functions indicate that *OCA2* is involved in supplying substrates for tyrosinase in melanin biosynthesis (Ancans *et al.,* 2001). Alternatively, *OCA2* may be involved in the intracellular pathway of tyrosinase during melanosome maturation (Toyofuku et al., 2002). Duffy et al. (2007) proposed that variations within the regulatory 5' control region are responsible for 90% of the variation in eye color pigmentation. Puri et al. (2000) also suggested that the *OCA2* gene product, a membrane melanosomal protein, can serve as an anion transporter, thereby helping to regulate the melanosomal pH.

Recent studies have revealed that the interaction between *OCA2* and *HERC2* may be responsible for determining the blue eye color in humans. In an association study with *OCA2* and *HERC2,* Branicki et al. (2009) verified possible interaction effects with other genes (*MC1R*, *ASIP*, and *SLC45A2*) also associated with variation in eye, hair and skin color in some populations. They concluded that the rs1800407 (Arg419Gln) polymorphism in the *OCA2* gene is associated with eye color and that there are significant interactions between *MC1R* and *HERC2* in determining skin and hair color in their Polish sample. Sturm (2009), on the other hand, found that variants in *OCA2* increase by 2.4 times the amount of melanin in melanocytes culture. In another study on a sample of individuals of Asian ancestry, Edwards et al. (2010) verified that the non-synonymous polymorphism His615Arg (rs1800414) of the *OCA2* gene is associated with skin pigmentation. Individuals with the G derived allele, which codes for the aminoacid arginine, showed less melanin index than those with the ancestral allele A, which encodes the aminoacid histidine. The methodology and analysis performed by Edwards et al. (2010) are consistent with other findings, also indicating that the evolution of light skin occurred at least in part independently in Europeans and Asians.

2.3. *SLC45A2 (MATP)*

The Membrane Associated Transporter Protein (*MATP*), known as *SLC45A2* (solute carrier family 45, member 2), has 7 exons and is located on chromosome 5p. The matp protein has 530 aminoacids (Sturm, 2009) and is considered a regulator of tyrosinase activity in human melanocytes (Smith et al., 2004). In the GWAS conducted by Nan et al. (2009a), they suggested that the rs16891982 in the *SLC45A2* gene is the most likely causal variant or is in strong linkage disequilibrium with the causal variant related to tanning ability.

Polymorphisms in the gene encoding the *Matp* protein proved to be related to variation in normal and abnormal skin color in Europeans. Lucotte et al. (2010) showed that the frequency of the Phe374Leu (rs16891982) polymorphism in Western Europe and North Africa may reflect the level of ultraviolet radiation received and the associated variation in skin color of individuals in these regions. This same polymorphism would also be associated with protection against the development of malignant melanoma (Guedj et al., 2008, Fernandez et al., 2008). Yuasa et al. (2006), in turn, suggested that the 374Phe allele may be an important causative factor in hypopigmentation in Caucasian populations. Cook et al. (2009) also analyzed the same polymorphism and found that the 374Leu allele was associated with high levels of tyrosinase and the homozygotes for it had lower levels of Matp transcripts in the European sample analyzed. Sturm (2009) observed that the these 100 homozygotes showed 2.6 times the amount of melanin in cultured melanocytes, as compared to the 374Phe homozygotes. The heterozygous genotype possessed an intermediate melanin index, suggesting an additive effect. Finally, in the genome-wide study conducted by Sabeti et al. (2007) it was found that the *SLC45A2* gene is probably under positive selection in Europeans, since the 374Phe allele is nearly fixed in this population, but absent in Asian and African samples.

2.4. *TYR*

Melanin synthesis also occurs with the participation of tyrosinase-related proteins (Huang et al., 2008). Tyrosinase has a critical role in cellular melanogenesis and therefore has been a target for skin pigmentation studies, including pharmacological tests involved in the care and esthetic maintenance of this pigmentation (An et al., 2009). The *TYR* gene, coding for tyrosinase, has five exons, is located on chromosome 11q, and produces a molecule of 529 aminoacids (Sturm, 2009).

TYR is one of the classic genes related to oculocutaneous albinism (Hutton & Spritz, 2008), and Meyler & Guldberg (2009) suggested that its product has a major effect on the hair and skin pigmentation. Bishop et al. (2009), on the other hand, related them to sun-sensitive skin, freckles, and a specific melanoma. An et al. (2009) used *TYR* small interfering RNA to test the effect of silencing this gene in some cell types. *TYR* silencing resulted in the suppression of melanin synthesis and decreased the viability of cells exposed to ultraviolet radiation.

In the GWAS conducted by Nan et al. (2009a), several SNPs were found significantly associated with tanning ability in a USA population of European ancestry, among them SNPs in the *TYR* gene. One of them (rs1393350) was the only which remained significant after adjusting for the analysis conducted. The SNP is also in strong linkage disequilibrium with

rs1126809 (Arg402Gln), a common and widely studied tyrosinase polymorphism. Tyrosinase is critical for melanosomal maturation, leading also to eumelanosome formation (Jimbow et al., 2000).

Fukai et al. (1995) and Morell et al., (1997) reported that one allele of the Arg402Gln polymorphism was correlated with reduced iris and retina pigmentation, due to low tyrosinase activity. In another study Nan et al. (2009b) also showed that this same allele was significantly related to tanning ability and skin color. Sturm (2009), in turn, showed that polymorphisms in other genes (*SLC45A2*, *SLC24A5* and *OCA2*) also cause variations in the tyrosinase activity. For example, in the *SLC24A5,* Ala111Thr polymorphism 111Ala/111Ala homozygotes showed 1.7x increased tyrosinase activity, compared to 111Thr/111Thr homozygotes. *SLC45A2* (Leu374Phe) and *OCA2* (rs12913832 T> C) polymorphisms were also correlated with tyrosinase activity. Homozygotes for *SLC45A2* 374Leu and *OCA2* TT independently presented a higher tyrosinase activity (2.8x and 2.1x, respectively), compared to 374PhePhe (*SLC45A2*) and CC (*OCA2*) genotypes.

2.5. *ASIP*

The gene encoding the agouti signaling protein (*ASIP*) has 3 exons and is located on chromosome 20q. Its product, together with other ligands, regulates the exchange between the synthesis of eumelanin and pheomelanin in melanocytes (Hida et al., 2009). It is known that *ASIP* regulates the pigmentation in mice, but its functional role in other animals and humans has not been well elucidated (Norris & Whan, 2008). The *ASIP* gene encodes a molecule with 132 aminoacids (Sturm, 2009), which coordinates melanocytes in the hair follicle to produce yellow or black pigment. The unregulated expression of agouti in mice carrying the dominant (yellow) allele is associated with pleiotropic effects including increased yellow fur pigmentation, obesity, diabetes and increased susceptibility to tumors (Perry et al., 1995).

Anno et al. (2008) attempted to identify alleles at multiple loci associated with differences in skin color in peoples of various origins using SNP genotyping and linkage disequilibrium. They analyzed 20 SNPs of several (*ASIP*, *TYR*, *OCA2*, *MC1R*, and others) genes in Canadians and Japanese and found significant differences between the two groups. Although the Asip protein was presented as a competitive antagonist of the *MC1R* gene product, the possibility of having other independent activity could not be ruled out, according to Hunt & Thody (1995). These authors showed that the rodent agouti protein causes suppression of melanin production in cultured melanoma cells. Moreover, this same protein decreased melanogenesis in human epidermal melanocytes in culture. Recent studies also suggest that the *ASIP* gene product regulates the expression of the *SLC24A5* gene (Nadeau et al., 2008).

Bonilla et al. (2005) showed that the 8818A allele acts as a strong antagonist to alpha-MSH, binding to Mc1r and then blocking the signaling mediated by cAMP in the melanocyte. The production of pheomelanin is thus favored, leading to the maturation of feomelanosomes and lighter coloring.

2.6. *SLC24A5 (NCKX5)*

One of the genes of the family of solute carriers (*SLC24A5*) encodes the *Nckx5* protein with 500 aminoacids, has 9 exons and is located on chromosome 15q (Sturm, 2009). Giardina et al., (2008) showed that the G (rs1426654 - Thr111Ala) allele together with two other intragenic markers (rs2555364 and rs16960620) vary in frequency among populations with different skin pigmentation.

Ginger et al. (2008) were the first to demonstrate that expression of the *SLC24A5* gene product is required for the production of melanin in differentiating human epidermal melanoblasts. It is believed that the *SLC24A5* gene encodes a calcium/sodium potassium dependent transporter present in the melanosomal membrane, and calcium may be required for the activation of a gene to produce Pmel17, required to form mature eumelanosome involved in eye color variation (Frudakis et al., 2003); melanosome pH regulation may also be a function of the SLC24A5 protein (Lamason et al., 2005). The role of pH in melanogenesis has been studied in detail, and there is evidence that pH changes affect tyrosinase maturation and its catalytic activity (Smith et al., 2004; Watabe et al., 2004).

SLC24A5, which is orthologous to the golden gene in zebrafish, has been widely cited in studies of human pigmentation. For example, in a study by Dimisianos et al. (2009) in Greece the rs1426654 (Thr111Ala) SNP was analyzed and the results were correlated with pigmentation traits and with *MC1R* genotypes. It was shown that the 111Thr allele was more prevalent among Greeks of darker skin, although this study shows a bias, since the majority of subjects (99%) were homozygous for this allele. *TYR*, *SLC45A2* and *SLC24A5* polymorphisms were studied in South Asian populations and their alleles may account for differences in skin reflectance (darker or lighter skin) through a simple additive model (Sturm, 2009). On the other hand, *SLC45A2/MATP*, *SLC24A5/NCKX5* and *OCA2/P* SNPs have been associated with natural variation of pigmentation traits in several human populations (Cook et al., 2009).

In a genome-scan for the detection and characterization of positive selection in human populations Sabeti et al. (2007) verified that *SLC24A5* and *SLC45A2* are subjected to this type of selection, while *EDAR* and *EDAR2R* were involved in the development of the hair follicle. It was also found that the *SLC24A5* gene Thr111Ala polymorphism showed the strongest signal of positive selection. Its ancestral allele (111Ala) predominates in Amerindians, Africans and East Asians. Conversely, 111Thr is almost fixed in Europeans and is also correlated with skin lighter pigmentation in admixed populations (Lamason et al., 2005). In a functional study conducted by Sturm (2009) it was found that the AlaAla genotype increases 2.2 times the amount of melanin in cultured melanocytes, compared to ThrThr. Heterozygous subjects possessed an intermediate level of melanin, suggesting an additive effect.

2.7. *ADAM17*

The gene encoding the Adam17 protein (*ADAM17* - a disintegrin and metalloprotease) has 19 exons and is located on chromosome 2p. Its product encodes 824 aminoacids and it is part of the Adam family of proteolytic enzymes, acting in the cleavage of other enzymes (Corrias et al., 2010; Yamamoto et al., 2010). It is also called *TACE* (Tumor necrosis factor-Alpha Converting Enzyme), and codifies a metalloprotease associated with the cell membrane

that is expressed in endothelial cells (Weskamp et al., 2010), adipocytes (Junyent et al., 2010) and is the primary enzyme responsible for catalyzing the release of membrane proteins anchored in the cell surface of Metazoan organisms (Caescu et al., 2009).

These metalloproteinases can rapidly modulate signaling events in the cell surface through proteolytic release of soluble forms of pro-ligands for cell receptors, and many regulatory pathways are affected by the *Adam17* activity; the detailed mechanisms of its activity, however, are not well understood (Willems et al., 2010). *Adam17* has pleiotropic effects and is therefore also associated with various diseases such as those leading to neurodegeneration, cancer and inflammation (Hoettecke et al., 2010, Long et al., 2010). Norton et al. (2007) have detected significant evidence of natural selection in *ADAM17* and *ATRN*. They found evidence of positive selection in these genes in Asians comparable to the *MATP* and *SLC24A5* signals found in Europeans.

2.8. *HERC2*

HERC2 is located upstream of *OCA2* (Eiberg et al., 2008). In three independent GWAS analyses the 15q13.1 region, which covers 38cM in the genome and includes genes *OCA2* and *HERC2*, was found to be the predominant region involved in human iris color (Kayser et al., 2008). Estimates indicated that 74% of the variance in human eye color could be explained by this region. The *HERC2* gene contains 93 exons, and the detailed function of its product is still unknown, but the gene encodes domains of conserved functional proteins involved in spermatogenesis, ubiquitin-mediated proteolysis and intracellular transport (Ji et al., 2000). Interactions between *OCA2* and *HERC2* through epistatic effects were observed (Kayser et al., 2008).

The melanin pigment within the iris is responsible for the human eye color (Sturm & Larsson, 2009). Fine mapping of this region identified a single base variation (rs12913832 - T/C) in *HERC2* intron 86, which explains much of the association with blue-brown eye color. Eiberg et al. (2008) had already identified two SNPs (rs12913832 and rs1129038) within the 15q13.1 region that were strongly associated with blue and brown eye colors. Kayser et al. (2008) also claimed that no other region on genes showed so consistent evidence of association with iris color. They also suggested that the testing of *OCA2-HERC2* markers could be used in forensic applications to predict eye color phenotypes of unknown persons of European origin. Partial *OCA2-HERC2* deletions are known in the Prader-Willi and Angelman syndromes, which are associated with oculocutaneous albinism or reduced pigmentation (Spritz et al., 1997).

So far the most consistent reports concerning eye color have been the association of three SNPs of *OCA2* intron 1 and two *HERC2* (rs129113832 and rs1129038) SNPs (Iida et al., 2009). Humbert (2008) suggested that blue eye color could be due to the change of an element located in *HERC2* suppressing *OCA2* expression.

2.9. *TPCN2*

The *TPCN2* (Two-Pore Channel segment 2) gene has 25 exons and is located on chromosome 11q. Its encoded protein has 752 aminoacids (Sturm, 2009). This protein

influences calcium transport, similarly to those of the *SLC24A4* (Sulem et al., 2007) and *SLC24A5* (Lamason et al., 2005) genes. Sulem et al. (2008) found that two coding polymorphisms (rs35264875 - Met484Leu and rs3829241 - Gly734Glu) could explain the difference between blond and brown hair individuals in Iceland.

2.10. Other Pigmentation Genes

The loci involved in pigmentation are related to a variety of physiological processes, including DNA damage responses, immunity and hypo/hyperpigmentation (Stinchcombe et al., 2004), melanocyte migration, melanogenesis control, hormone responses (MSH-like and sex hormones or vitamin D), or prostaglandin synthesis. The latter may mediate changes in post-inflammatory pigmentation by modulating melanocyte dendricity and melanin synthesis (Scott et al., 2004). These loci include *ADAMTS20*, *AP3D1*, *AP3M2*, *AR*, *ASMT*, *ATRN*, *BLOC1S3*, *BRCA1*, *CHS1* (*LYST*), *CNO*, *CYP24A1*, *CYP27B1*, *DC6*, *DCT*, *DDC*, *EDA*, *EDN3*, *EDNRB*, *ERCC2*, *ESR1*, *ESR2*, *GGT1*, *GPR143*, *HPS1*, *HPS2* (*AP3B1*), *HPS3*, *HPS5*, *HPS6*, *HPS7* (*DTNBP1*), *IKBKG*, *IRF4*, *KIT*, *KITLG*, *MAGMAS*, *MAMAL1*, *MDM2*, *MGRN1*, *MITF*, *MLANA*, *MLPH*, *MTNR1A*, *MTNR1B*, *MUTED*, *MYO5A*, *MYO7A*, *OSTM1*, *PAR2*, *PAX3*, *PER1*, *PGR*, *PKCb*, *PLDN*, *POMC*, *PTGER1*, *PTGER2*, *PTGER3*, *PTGFR*, *PTGIR*, *RAB27A*, *RABGGTA*, *RAD50*, *RXRA*, *SILV* (*PM17*), *SOX10*, *TBX2*, *TH*, *TP53*, *TP53BP1*, *TYRP1*, *TYRP2* (*DCT*), *VDR*, *VDRIP*, *VPS18*, *VPS33A*, *YARS*, and *ZNFN1A1*, besides those already mentioned above.

Many of these genes show no robust evidence for positive selection (Izagirre et al., 2006; Myles et al., 2007; Norton et al., 2007), differently from those previously mentioned. But this is expected due to the complex nature of the phenotypes considered.

3. Forensic Application and Perspectives

The initial attempt to apply human DNA analysis to forensic investigations was based in DNA "fingerprints", individual-specific genetic profiles defined by a set of hypervariable markers' genotypes (Jeffreys et al., 1985). These "fingerprints" could be analyzed with southern-blot hybridization and used for human identification and parenthood testing. Since then, however, the repertoire of genetic markers used in forensic routine has grown substantially and several advances in molecular biology had brought much progress to the forensic sciences.

One of such advances is the possibility to infer the person's ethnicity and/or physical appearance by using genetic markers. Ancestry Informative Markers (AIMs), genetic loci showing alleles with large frequency differences between continental populations (Shriver et al., 2003), have been employed for this purpose. These markers enhance the power of an already effective system of DNA analysis based on microsatellite variation for the identification of unmatched samples in, for instance, cases such as mass-disasters due to terrorist attacks. It is believed that these markers could also provide information about the general appearance of a person (Budowle and Van Daal, 2008). But although they could

provide some lead in certain populations, AIMs may not be completely associated to human physical appearance in admixed populations (Parra et al., 2003).

In cases where police investigations fail to identify the attackers, and especially when the biological evidence found at crime scenes do not match those existing in the DNA police database, it is expected the development in the near future of predictions of externally visible features (Externally Visible Characteristics - EVCs) based on DNA (Kayser & Schneider, 2009). Physical features that are explained by a small number of loci and few alleles, easy to detect using standard methods of linkage analysis and genetic association, are the most promising.

Usually the first step in analyzing a crime-scene sample is to produce a profile of short tandem repeats (STRs) using multiplex amplification kits. When this profile do not match those of a police database, any additional information is extremely valuable (Werrett, 2005), including tests for predicting specific phenotypes (Tully, 2007). SNPs on the Y chromosome (Y-SNPs) are usually investigated for their potential to predict the geographic origin of a given male suspect. Unfortunately, due to the presence of admixed individuals, it is occasionally difficult to predict the geographic ancestry of an individual using Y-SNPs only. In a study conducted by Sims & Ballantyne (2008), using the *SLC24A5* gene variants in combination with Y-SNPs and other lineage markers, the authors were able to blindly differentiate Africans, Europeans and Asians. The eye color prediction could also be of great value in certain crime investigations. Mengel-From et al., (2010), for instance, developed predictive values for various SNP combinations in *HERC2*, *OCA2*, and *SLC45A2* (*MATP*) genes for blue/brown eye color.

Grimes et al., (2001) described a multiplex minisequencing protocol to quick scan DNA samples for the presence of 12 variants of the *MC1R* gene. In another study Frudakis et al. (2003) tested 754 SNPs in 851 individuals, and identified 61 SNPs associated with iris color. Nearly half of these SNPs were independently associated with this characteristic, while the remainder was associated only if considered as part of a haplotype or diplotype. Some companies already provide multiplex PCR-based tests for predicting iris color, and based on the literature, it is estimated that the inferences are correct in 92% of the cases.

With regard to ethical issues, the forensic scientific community recommends to avoid typing markers that are predictive of disease susceptibility, to protect the privacy of the persons involved (Tully, 2007).

4. Evolutionary Background

The genetic basis underlying the normal variation in skin, hair and eye pigmentation has been the subject of intense research to verify the presence of any adaptive component. A combination of approaches has been used, including comparative genomics of candidate genes, identification of regions of the human genome under positive selection, allele-specific association studies, and genome-wide scans (Sturm, 2009).

Although the combination of fair skin and intense sun exposure results in a significant individual health risk, it is unclear whether this risk has evolutionary significance at the population level. Therefore, several adaptive hypotheses have been proposed to explain the evolution of human skin pigmentation, such as photo-protection against sun-induced cancer,

sexual selection, vitamin D synthesis or photo-protection from photo-unstable compounds. It is expected that if skin pigmentation is adaptive, we could detect signs of the action of positive selection in some of the genes involved (Izagirre et al., 2006).

Several regions of the human genome have recently been identified as subjected to the action of positive selection. The combination of new methods, including large SNP banks at the population level, as well as new and robust analytic methodologies show clear evidence for selection acting on several pigmentation genes in different populations (Voight et al., 2006; Norton et al., 2007; Sabeti et al., 2007; Williamson et al., 2007; Johansson & Gyllensten, 2008). These studies also showed that the process probably occurred many times in multiple geographic locations around the globe, in ancient and recent human populations.

Jablonski & Chaplin (2000) and Chaplin (2004), for example, have confirmed a strong positive correlation between skin pigmentation and intensity of ultraviolet radiation. However, although there is strong evidence that the variation in pigmentation have been influenced by natural selection, little is known about how selection has affected the genetic architecture of pigmentation loci in different populations, even when these populations have experienced similar levels of ultraviolet radiation in their evolutionary trajectory. The dark skin that characterizes many African populations south of the Sahara and Melanesians may be due to shared ancestral variants or new genetic adaptations. Similarly, the light skin of Europeans and East Asians may have a common genetic origin or can be the result of independent adaptations to environments of low ultraviolet radiation.

Norton et al. (2007) described the allele frequencies of polymorphisms in the *ASIP* (A8818G), *OCA2* (A355G), *TYR* (Ser192Tyr), *MATP* (Phe374Leu), and *SLC24A5* (Thr111Ala) in various populations. They wanted to verify the presence of positive selection for these polymorphisms in Africans, Asians, Europeans and a small group of Native Americans. An evolutionary model for the independent selection of variants of pigmentation genes in East Asian, European and West African populations was proposed, leading to the conclusion of convergent evolution in the case of lighter colored skin (McEvoy et al., 2006; Norton et al., 2007). In Europeans and Asians, signs of positive selection were confirmed for *SLC24A5* (Lamason et al., 2005; Norton et al., 2007) and *SLC45A2* (Soejima et al., 2006; Norton et al., 2007) gene variants, which result in lighter skin. Norton et al. (2007) also concluded that light pigmentation in Europeans is due, at least in part, to the effects of sexual selection and/or positive directional selection, and not simply by relaxation of functional constraints, as previously suggested.

Izagirre et al., (2006) and Myles et al., (2007) used public data banks (HapMap and Pelergen), with African, African-American, European and East Asian genotypic information and performed an F_{st} analysis. Under a neutral model, F_{st} is determined by the demographic history (basically gene flow and drift) of the groups considered and would affect all loci in a similar way. In contrast, under the action of selection, specific loci would show different patterns. Balancing or purifying selection tend to reduce F_{st} values while positive selection would lead to increased values when compared to those of other genomic regions. The two studies identified signs of selection in *SLC24A5* and *SLC45A2* (*MATP*) in European and East Asian individuals. In contrast, no signs of selection were found for *MC1R*.

5. Improving Pigmentation Phenotypic Data For Forensic and General Studies

In studies that seek to relate genotype-phenotype in pigmentation characteristics, it is recommended that these characteristics should be examined as objectively as possible, avoiding qualitative ratings for skin pigmentation such as 'black', 'dark brown', 'light brown' or 'white', since classification criteria may be diverse in different countries and ethnicities.

It is suggested that frontal photographs should be made of the study participants, to make a data log archive of eyes and face pigmentation. Photographs of the back of the head can also be made to obtain data from hair color. Quantitative measurements of pigmentation through the contents of melanin in exposed/non-exposed regions, as described by Stokowski et al. (2007), are recommended. These measures could be made by reflectance spectrophotometry, for instance, DermaSpectrometer, Cortex Technology, Hadsund, Denmark. This instrument emits light at the green (568nm) and red (655nm) regions of the visible spectrum and the photo-detector measures the amount of light reflected by the skin. These measures are used to estimate the melanin content of skin, which is expressed as a Melanin Index (M). In human populations M ranges from about 20 (people with lighter skin) to near 100 (individuals with darker skin). For more accurate readings, these measures should be carried out preferably indoors and in a well-lighted room. More information about the DermaSpectrometer can be found in Shriver & Parra (2000).

A traditional method for hair classification involves its observation under a microscope considering such characteristics as color, length, diameter, pigment granules and other morphological traits (Saferstein, 2001; Bednarek, 2004; SWGMAT, 2005). Reflectance spectrophotometry is also used to determine hair color, which is more consistent than digital image analysis (Vaughn et al., 2008, 2009b). The CIE L*a*b* apparatus is used both for digital image analysis by microscopy, or for measuring macroscopic color (Vaughn et al., 2009a). The CIE L*a*b* model was developed by the *Commission Internationale de l'Eclairage* and measures the color in three axes, that correspond to the trichromatic human perception and reflects the degree of change in color that humans can perceive (Ford & Roberts, 1998). In this system, the brightness or intensity of a color is measured in the L* axis on a scale from 0 (black) to 100 (white). Color, in turn, is measured first along the a* axis, varying from -100 (green) to +100 (red); and then in the *b** axis, varying from -100 (blue) to +100 (yellow). The L*a*b* unit is regarded as the minimum difference that the human eye can observe (TASI, 2004).

Iris color categorization has been carried out using a variety of descriptive terms for self-assessment, as well as the use of a trained observer to describe, or qualitatively classify it using photographic standards (Seddon et al., 1990; Franssen et al., 2008). None of these methods is completely reliable and thus automated photographic methods have been developed to improve the accuracy in classification (Takamoto et al., 2001; Niggemann et al., 2003). A method which uses a digital camera and a computer software to analyze the light reflected from the iris was described in detail, with the sum of a range of color and brightness values used to derive an expression called the Iris Melanin Index (IMI), suitable for categorizing a population survey, but possibly still missing details of the iris pattern (Frudakis, 2008).

The most important point to emphasize is that objective, not subjective classifications should be used. Qualitative subdivision of continuous traits also oversimplify their nature (Liu et al., 2010). If indeed it is not possible to quantitatively measure hair and eye colors, care must be taken that the classification is made in accordance with previous studies, with the help of color standards. The most common hair characterizations involve 'black', 'brown' (without subdivisions, to avoid a subjective qualification), 'red', 'blond' or 'white'; while eye pigmentation may be characterized as 'black', 'brown', 'blue', 'green ' or 'grey'; whenever possible, the presence of brown spots around the pupil or any other eye brown spot should also be recorded (Eiberg et al., 2008).

6. REFERENCES

Albert, DM; Green, WR; Zimbric, ML; Lo, C; Gangnon, RE; Hope, KL; Gleiser, J. Iris melanocyte numbers in Asian, African American, and Caucasian irides. *Trans. Am. Ophthalmol. Soc*. 101: 217–222. 2003.

An, SM; Koh, JS; Boo, YC. Inhibition of melanogenesis by tyrosinase siRNA in human melanocytes. *BMB Rep.* 42(3):178-183. 2009.

Ancans, J; Tobin, DJ; Hoogduijn, MJ; Smit, NP; Wakamatsu, K; Thody, AJ. Melanosomal pH controls rate of melanogenesis, eumelanin/ phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. *Exp Cell Res.* 268: 26–35. 2001.

Anno, S; Abe, T; Yamamoto, T. Interactions between SNP alleles at multiple loci contribute to skin color differences between caucasoid and mongoloid subjects. *Int J Biol Sci.* 4(2):81-86. 2008.

Barsh, GS. Regulation of pigment type switching by agouti, melanocortin signaling, attractin and mahoganoid. In *The Pigmentary System,* 2nd edn J.J. Nordlund, R.E. Boissy, V.J. Hearing, R.A. King, W.S. Oetting, and J.P. Ortonne, eds (Massachusetts: Blackwell Publishing), pp. 395–409. 2006.

Beaumont, KA; Newton, RA; Smit, DJ; Leonard, JH; Stow, JL; Sturm, RA. Altered cell surface expression of human MC1R variant receptor alleles associated with red hair and skin cancer risk. *Hum Mol Genet.* 14: 2145–2154. 2005.

Beaumont, KA; Shekar, SN; Newton, RA; James, MR; Stow, JL; Duffy, DL; Sturm, RA. Receptor function, dominant negative activity and phenotype correlations for MC1R variant alleles. *Hum Mol Genet.* 16: 2249–2260. 2007.

Bednarek, J. An attempt to establish objective criteria for morphological examinations of hairs using the image analysis system. *Probl Forensic Sci* 56: 65–77. 2004.

Blum, H. Does the melanin pigment of human skin have adaptive value? *Q Rev Biol*. 36:50–63. 1961.

Bonilla, C; Boxill, LA; Donald, SA; Williams, T; Sylvester, N; Parra, EJ; Dios, S; Norton, HL; Shriver, MD; Kittles, RA. The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. *Hum Genet.* 116 (5): 402–406. 2005.

Box, N; Wyeth, J; O'Gorman, L; Martin, NG; Sturm, RA. Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. *Hum Mol Genet*. 6 (11): 1891–1897. 1997.

Brace, C. Structural reduction in evolution. *Am Nat.* 97:39–49. 1963.

Branicki, W; Brudnik, U; Wojas-Pelc, A. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. *Ann Hum Genet.* 73(2):160-170. 2009.

Budowle, B & Van Daal, A. Forensically relevant SNP classes. *BioTechniques*. 44 (5): 603-610 (25th Anniversary Issue, April 2008). 2008.

Buscà, R & Ballotti, R. Cyclic AMP a key messenger in the regulation of skin pigmentation. *Pigment Cell Res.* 13(2): 60–69. 2000.

Caescu, CI; Jeschke, GR; Turk, BE. Active-site determinants of substrate recognition by the metalloproteinases TACE and ADAM10. *Biochem J.* 424(1):79-88. 2009.

Chaplin, G. Geographic distribution of environmental factors influencing human skin coloration. *Am J Phys Anthropol.* 125:292–302. 2004.

Cheli, Y; Luciani, F; Khaled, M; Beuret, L; Bille, K; Gounon, P; Ortonne, JP; Bertolotto, C; Ballotti, R. {alpha}MSH and Cyclic AMP elevating agents control melanosome pH through a protein kinase A-independent mechanism. *J Biol Chem.* 284(28):18699-18706. 2009.

Cook, AL; Chen, W; Thurber, AE; Smit, DJ; Smith, AG; Bladen, TG; Brown, DL; Duffy, DL; Pastorino, L; Bianchi-Scarra, G; Leonard, JH; Stow, JL; Sturm, RA. Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. *J Invest Dermatol.* 129(2):392-405. 2009.

Coplan, RJ; Coleman, B; Rubin, KH. Shyness and little boy blue: iris pigmentation, gender, and social wariness in preschoolers. *Dev. Psychobiol.* 32: 37–44. 1998.

Corrias, MV; Gambini, C; Gregorio, A; Croce, M; Barisione, G; Cossu, C; Rossello, A; Ferrini, S; Fabbi, M. Different subcellular localization of ALCAM molecules in neuroblastoma: association with relapse. *Cell Oncol.* 32(1-2):77-86. 2010.

Darwin, C. The descent of man. Princeton University Press, Princeton. 1871.

Da Costa, EA; Castro, JC; Macedo, ME. Iris pigmentation and susceptibility to noise-induced hearing loss. *Int. J. Audiol.* 47: 115–118. 2008.

Dimisianos, G; Stefanaki, I; Nicolaou, V; Sypsa, V; Antoniou, C; Poulou, M; Papadopoulos, O; Gogas, H; Kanavakis, E; Nicolaidou, E; Katsambas, AD; Stratigos, AJ. A study of a single variant allele (rs1426654) of the pigmentation-related gene SLC24A5 in Greek subjects. *Exp Dermatol.* 18(2):175-177. 2009.

Duffy, DL; Box, NF; Chen, W; Palmer, JS; Montgomery, GW; James, MR; Hayward, NK; Martin, NG; Sturm, RA. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum Mol Genet.* 13: 447–461. 2004.

Duffy, DL; Montgomery, GW; Chen, W; Zhao, ZZ; Le, L; James, MR; Hayward, NK; Martin, NG; Sturm, RA. A three single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am J Hum Genet.* 80: 241–252. 2007.

Edwards, M; Bigham, A; Tan, J; Li, S; Gozdzik, A; Gozdzik, A; Ross, K; Jin, L; Parra, EJ. Association of the OCA2 Polymorphism His615Arg with melanin content in East Asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genet.* 6(3): e1000867. 2010.

Eiberg, H; Troelsen, J; Nielsen, M; Mikkelsen, A; Mengel-From, J; Kjaer, KW; Hansen, L. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet.* 123(2):177-187. 2008.

Enserink, M. Can this DNA sleuth help catch criminals? *Science.* 331: 838-840. 2011.

Fernandez, LP; Milne, RL; Pita, G; Avilés, JA; Lázaro, P; Benítez, J; Ribas, G. SLC45A2: a novel malignant melanoma-associated gene. *Hum Mutat*. 29(9):1161-1167. 2008.

Ford, A & Roberts, A. *Colour space conversions*. Available at www.poynton.com/PDFs/coloureq.pdf, accessed February 1, 2007. 1998.

Franssen, J; Coppens, JE; Van Den Berg, TJTP. Grading of iris color with an extended photographic reference set. *J Optom*. 1: 36–40. 2008.

Frost, P. European hair and eye color: a case of frequency dependent sexual selection? *Evol Hum Behav*. 27: 85–103. 2006.

Frudakis, TN. *Molecular Photofitting – Predicting Ancestry and Phenotype Using DNA* (Burlington, MA, USA: Academic Press). 2008.

Frudakis, T; Thomas, M; Gaskin, Z; Venkateswarlu, K; Chandra, KS; Ginjupalli, S; Gunturi, S; Natrajan, S; Ponnuswamy, VK; Ponnuswamy, KN. Sequences associated with human iris pigmentation. *Genetics*. 165(4): 2071–2083. 2003.

Fukai, K; Holmes, SA; Lucchese, NJ; Siu, VM; Weleber, RG; Schnur, RE; Spritz, RA. Autosomal recessive ocular albinism associated with a functionally significant tyrosinase gene polymorphism. *Nat Genet*. 9:92–95. 1995.

Gantz, I; Yamada, T; Tashiro, T; Konda, Y; Shimoto, Y; Miwa, H; Trent, JM. Mapping of the gene encoding the melanocortin-1 (alpha-melanocyte stimulating hormone) receptor (MC1R) to human chromosome 16q24.3 by fluorescence in situ hybridization. *Genomics*. 19 (2): 394–395. 1994.

Giardina, E; Pietrangeli, I; Martínez-Labarga, C; Martone, C; de Angelis, F; Spinella, A; De Stefano, G; Rickards, O; Novelli, G. Haplotypes in SLC24A5 gene as ancestry informative markers in different populations. *Curr Genomics*. 9(2):110-114. 2008.

Ginger, RS; Askew, SE; Ogborne, RM; Wilson, S; Ferdinando, D; Dadd, T; Smith, AM; Kazi, S; Szerencsei, RT; Winkfein, RJ; Schnetkamp, PPM; Green, MR. SLC24A5 encodes a trans-Golgi network protein with potassium-dependent sodium-calcium exchange activity that regulates human epidermal melanogenesis. *J Biol Chem*. 283 (9): 5486–5495. 2008.

Goding, CR. Mitf from neural crest to melanoma: signal transduction and transcription in the melanocyte lineage. *Genes Dev*. 14(14): 1712–1728. 2000.

Goel, N; Terman, M; Terman, JS. Depressive symptomatology differentiates subgroups of patients with seasonal affective disorder. *Depress Anxiety* 15, 34–41. 2002.

Grimes, EA; Noake, PJ; Dixon, L; Urquhart, A. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci Int*. 122(2-3): 124–129. 2001.

Guedj, M; Bourillon, A; Combadières, C; Rodero, M; Dieudé, P; Descamps, V; Dupin, N; Wolkenstein, P; Aegerter, P; Lebbe, C; Basset-Seguin,N; Prum, B; Saiag, P; Grandchamp, B; Soufir, N; MelanCohort Investigators. Variants of the MATP/SLC45A2 gene are protective for melanoma in the French population. *Hum Mutat*. 29(9):1154-1160. 2008.

Hida, T; Wakamatsu, K; Sviderskaya, EV; Donkin, AJ; Montoliu, L; Lynn Lamoreux, M; Yu, B; Millhauser, GL; Ito, S; Barsh, GS; Jimbow, K; Bennett, DC. Agouti protein, mahogunin, and attractin in pheomelanogenesis and melanoblast-like alteration of melanocytes: a cAMP-independent pathway. *Pigment Cell Melanoma Res*. 22(5): 623-634. 2009.

Hoettecke, N; Ludwig, A; Foro, S; Schmidt, B. Improved synthesis of ADAM10 inhibitor GI254023X. *Neurodegener Dis.* 7(4):232-238. 2010.

Huang, YH; Lee, TH; Chan, KJ; Hsu, FL; Wu, YC; Lee, MH. Anemonin is a natural bioactive compound that can regulate tyrosinase-related proteins and mRNA in human melanocytes. *J Dermatol Sci.* 49(2):115-123. 2008.

Humbert, P. Quoi de neuf en recherche dermatologique? *Ann Dermatol Venereol.* 135 Suppl 7:S326-334. 2008.

Hunt, G & Thody, AJ. Agouti protein can act independently of melanocyte-stimulating hormone to inhibit melanogenesis. *J Endocrinol.* 147(2):R1-R4. 1995.

Hutton, SM & Spritz, RA. A comprehensive genetic study of autosomal recessive ocular albinism in Caucasian patients. *Invest Ophthalmol Vis Sci.* 49(3):868-872. 2008.

Iida, R; Ueki, M; Takeshita, H; Fujihara, J; Nakajima, T; Kominato, Y; Nagao, M; Yasuda, T. Genotyping of five single nucleotide polymorphisms in the OCA2 and HERC2 genes associated with blue-brown eye color in the Japanese population. *Cell Biochem Funct.* 27(5):323-327. 2009.

Imesch, PD; Wallow, IH; Albert, DM. The color of the human eye: a review of morphologic correlates and of some conditions that affect iridial pigmentation. *Surv Ophthalmol.* 41(Suppl 2): S117–S123. 1997.

Ito, S. The IFPCS presidential lecture: a chemist's view of melanogenesis. *Pigment Cell Res.* 16(3): 230–236. 2003.

Izagirre, N; García, I; Junquera, C; de la Rúa, C; Alonso, S. A scan for signatures of positive selection in candidate loci for skin pigmentation in humans. *Mol Biol Evol.* 23(9):1697-1706. 2006.

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

Jeffreys, AJ; Wilson, V; Thein, SL. Hypervariable minisatellite regions in human DNA. *Nature.* 314:67-73. 1985.

Ji, Y; Rebert, NA; Joslin, JM; Higgins, MJ; Schultz, RA; Nicholls, RD. Structure of the highly conserved HERC2 gene and of multiple partially duplicated paralogs in human. *Genome Res.* 10(3): 319–329. 2000.

Jimbow, K; Hua, C; Gomez, PF; Hirosaki, K; Shinoda, K; Salopek, TG; Matsusaka, H; Jin, HY; Yamashita, T. Intracellular vesicular trafficking of tyrosinase gene family protein in eu- and pheomelanosome biogenesis. *Pigment Cell Res.* 13(Suppl 8):110–117. 2000.

Johansson, A & Gyllensten, U. Identification of local selective sweeps in human populations since the exodus from Africa. *Hereditas.* 145: 126–137. 2008.

Junyent, M; Parnell, LD; Lai, CQ; Arnett, DK; Tsai, MY; Kabagambe, EK; Straka, RJ; Province, M; An, P; Smith, CE; Lee, YC; Borecki, I; Ordovás, JM. ADAM17_i33708A>G polymorphism interacts with dietary n-6 polyunsaturated fatty acids to modulate obesity risk in the Genetics of Lipid Lowering Drugs and Diet Network study. *Nutr Metab Cardiovasc Dis.* 20(10): 698-705. 2010.

Kayser, M & Schneider, PM. DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. *FSI Genetics.* 3: 154–161. 2009.

Kayser, M; Liu, F; Janssens, AC; Rivadeneira, F; Lao, O; Van Duijn, K; Vermeulen, M; Arp, P; Jhamai, MM; Van Ijcken, WF; Den Dunnen, JT; Heath, S; Zelenika, D; Despriet, DD; Klaver, CC; Vingerling, JR; De Jong, PT; Hofman, A; Aulchenko, YS; Uitterlinden, AG;

Oostra, BA; Van Duijn, CM. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet.* 82(2):411-423. 2008.

King, R; Willaert, R; Schmidt, R; Pietsch, J; Savage, S; Brott, MJ; Fryer, JP; Summers, CG; Oetting, WS. MC1R mutations modify the classic phenotype of oculocutaneous albinism type 2 (OCA2). *Am J Hum Genet.* 73 (3): 638–645. 2003.

Kobayashi, H & Kohshima, S. Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J Hum Evol.* 40: 419–435. 2001.

Kollias, N; Sayer, R; Zeise, L; Chedekel, M. Photoprotection by melanin. *J Photochem Photobiol* B. 9:135–160. 1991.

Lamason, RL; Mohideen, MA; Mest, JR; Wong, AC; Norton, HL; Aros, MC; Jurynec, MJ; Mao, X; Humphreville, VR; Humbert, JE; Sinha, S; Moore, JL; Jagadeeswaran, P; Zhao, W; Ning, G; Makalowska, I; McKeigue, PM; O'donnell, D; Kittles, R; Parra, EJ; Mangini, NJ; Grunwald, DJ; Shriver, MD; Canfield, VA; Cheng, KC. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science.* 310(5755):1782-1786. 2005.

Lamoreux, ML; Zhou, BK; Rosemblat, S; Orlow, SJ. The pinkeyed-dilution protein and the eumelanin/pheomelanin switch: in support of a unifying hypothesis. *Pigment Cell Res.* 8: 263–270. 1995.

Lee, ST; Nicholls, RD; Jong, MT; Fukai, K; Spritz, RA. Organization and sequence of the human P gene and identification of a new family of transport proteins. *Genomics.* 26(2): 354–363. 1995.

Liu, F; Wollstein, A; Hysi, PG; Ankra-Badu, GA; Spector, TD; Park, D; Zhu, G; Larsson, M; Duffy, DL; Montgomery, GW; Mackey, DA; Walsh, S; Lao, O; Hofman, A; Rivadeneira, F; Vingerling, JR; Uitterlinden, AG; Martin, NG; Hammond, CJ; Kayser, M. Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet.* 6(5):e1000934. 2010.

Long, C; Wang, Y; Herrera, AH; Horiuchi Dagger, K; Walcheck, B. In vivo role of leukocyte ADAM17 in the inflammatory and host responses during E. coli-mediated peritonitis. *J Leukoc Biol.* 87(6):1097-1101. 2010.

Lucotte, G; Mercier, G; Diéterlen, F; Yuasa, I. A decreasing gradient of *374F* allele frequencies in the skin pigmentation gene SLC45A2, from the north of West Europe to North Africa. *Biochem Genet.* 48(1-2):26-33. 2010.

Makova, K & Norton, H. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. *Peptides.* 26: 1901–1908. 2005.

Marks, MS & Seabra, MC. The melanosome: membrane dynamics in black and white. *Nat Rev Mol Cell Biol.* 2(10): 738–748. 2001.

McEvoy, B; Beleza, S; Shriver, MD. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Hum Mol Genet.* 15 (Spec no. 2): R176–R181. 2006.

Mengel-From, J; Børsting, C; Sanchez, JJ; Eiberg, H; Morling, N. Human eye colour and HERC2, OCA2 and MATP. *Forensic Sci Int Genet.* 4(5): 323-328. 2010.

Morell, R; Spritz, RA; Ho, L; Pierpont, J; Guo, W; Friedman, TB; Asher, JH Jr. Apparent digenic inheritance of Waardenburg syndrome type 2 (WS2) and autosomal recessive ocular albinism (AROA). *Hum Mol Genet.* 6(5):659-664. 1997.

Myles, S; Somel, M; Tang, K; Kelso, J; Stoneking, M. Identifying genes underlying skin pigmentation differences among human populations. *Hum Genet.* 120: 613–621. 2007.

Nadeau, NJ; Minvielle, F; Ito, S; Inoue-Murayama, M; Gourichon, D; Follett, SA; Burke, T; Mundy, NI. Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene. *Genetics.* 178(2):777-786. 2008.

Nan, H; Kraft, P; Qureshi, AA; Guo, Q; Chen, C; Hankinson, SE; Hu, FB; Thomas, G; Hoover, RN; Chanock, S; Hunter, DJ; Han, J. Genome-wide association study of tanning phenotype in a population of European ancestry. *J Invest Dermatol.* 129: 2250–2257. 2009a.

Nan, H; Kraft, P; Hunter, DJ; Han, J. Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. *Int J Cancer.* 125(4):909-917. 2009b.

Nei, M. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci USA.* 104 (30): 12235-12242. 2007.

Newton, JM; Cohen-Barak, O; Hagiwara, N; Gardner, JM; Davisson, MT; King, RA; Brilliant, MH. Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4. *Am J Hum Genet.* 69(5): 981–988. 2001.

Niggemann, B; Weinbauer, G; Vogel, F; Korte, R. A standardized approach for iris color determination. *Int J Toxicol.* 22, 49–51. 2003.

Norris, BJ & Whan, VA. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res.* 18(8):1282-1293. 2008.

Norton, HL; Kittles, RA; Parra, E; McKeigue, P; Mao, X; Cheng, K; Canfield, VA; Bradley, DG; McEvoy, B; Shriver, MD. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol Biol Evol.* 24(3):710-722. 2007.

Parra, FC; Amado, RC; Lambertucci, JR; Rocha, J; Antunes, CM; Pena, SDJ. Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci USA.* 100:177–182. 2003.

Perry, WL; Hustad, CM; Swing, DA; Jenkins, NA; Copeland, NG. A transgenic mouse assay for agouti protein activity. *Genetics.* 140(1):267-274. 1995.

Pho, LN & Leachman, SA. Genetics of pigmentation and melanoma predisposition. *G Ital Dermatol Venereol.* 145(1):37-45. 2010.

Prota, G; Hu, DN; Vincensi, MR; Mccormick, SA; Napolitano, A. Characterization of melanins in human irides and cultured uveal melanocytes from eyes of different colors. *Exp Eye Res.* 67, 293–299. 1998.

Puri, N; Gardner, JM; Brilliant, MH. Aberrant pH of melanosomes in pink-eyed dilution (P) mutant melanocytes. *J Invest Dermatol.* 115: 607–613. 2000.

Rana, B; Hewett-Emmett, D; Jin, L; Chang, BH; Sambuughin, N; Lin, M; Watkins, S; Bamshad, M; Jorde, LB; Ramsay, M; Jenkins, T; Li, WH. High polymorphism at the human Melanocortin 1 Receptor locus. *Genetics.* 151 (4): 1547–1557. 1999.

Rebbeck, TR; Kanetsky, PA; Walker, AH; Holmes, R; Halpern, AC; Schuchter, LM; Elder, DE; Guerry, D. P gene as an inherited biomarker of human eye color. *Cancer Epidemiol Biomarkers Prev.* 11: 782–784. 2002.

Relethford, JH. Hemispheric difference in human skin color. *Am J Phys Anthropol.* 104: 449–457. 1997.

Rogers, A; Iltis, D; Wooding, S. Genetic variation at the MC1R locus and time since loss of human body hair. *Curr Anthropol.* 45:105–108. 2004.

Rowe, PJ & Evans, P. Ball color, eye color, and a reactive motor skill. *Percept Mot Skills*. 79: 671–674. 1994.

Sabeti, PC; Varilly, P; Fry, B; Lohmueller, J; Hostetter, E; Cotsapas, C; Xie, X; Byrne, EH; McCarroll, SA; Gaudet, R; Schaffner, SF; Lander, ES; International HapMap Consortium. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 449(7164):913-918. 2007.

Saferstein, R. *Criminalistics*. Upper Saddle River, NJ: Prentice Hall. 2001.

Schioth, HB; Ohillips, SR; Rudzish, R; Birch-Mahin, MA; Wikberg, JE; Rees, JL. Loss of function mutations of the human melanocortin 1 receptor are common and associated with red hair. *Biochem Biophys Res Commun*. 260: 488–491. 1999.

Scott, G; Leopardi, S; Printup, S; Malhi, N; Seiberg, M; Lapoint, R. Proteinase-activated receptor-2 stimulates prostaglandin production in keratinocytes: analysis of prostaglandin receptors on human melanocytes and effects of PGE2 and PGF2alpha on melanocyte dendricity. *J Investig Dermatol*. 122: 1214–1224. 2004.

Seddon, JM; Sahagian, CR; Glynn, RJ; Sperduto, RD; Gragoudas, ES. Evaluation of an iris color classification system. The Eye Disorders Case-control Study Group. *Invest Ophthalmol Vis Sci*. 31: 1592–1598. 1990.

Shriver, MD & Parra, EJ. Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. *Am J Phys Anthropol*. 112: 17–27. 2000.

Shriver, MD; Parra, EJ; Dios, S; Bonilla, C; Norton, H; Jovel, C; Pfaff, C; Jones, C; Massac, A; Cameron, N; Baron, A; Jackson, T; Argyropoulos, G; Jin, L; Hoggart, CJ; McKeigue, PM; Kittles RA. Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet*. 112 (4): 387–399. 2003.

Sims, LM & Ballantyne J. The golden gene (SLC24A5) differentiates US sub-populations within the ethnically admixed Y-SNP haplogroups. *Leg Med (Tokyo)*. 10(2):72-77. 2008.

Smith, DR; Spaulding, DT; Glenn, HM; Fuller, BB. The relationship between Na(+)/H(+) exchanger expression and tyrosinase activity in human melanocytes. *Exp Cell Res*. 298:521–534. 2004.

Soejima, M; Tachida, H; Ishida, T; Sano, A; Koda, Y. Evidence for recent positive selection at the human AIM1 locus in a European population. *Mol Biol Evol*. 23: 179–188. 2006.

Spritz, RA; Bailin, T; Nicholls, RD; Lee, ST; Park, SK; Mascari, MJ; Butler, MG. Hypopigmentation in the Prader-Willi syndrome correlates with P gene deletion but not with haplotype of the hemizygous P allele. *Am J Med Genet*. 71: 57–62. 1997.

Stinchcombe, J; Bossi, G; Griffiths, CM. Linking albinism and immunity: the secrets of secretory lysosomes. *Science*. 305: 55–59. 2004.

Stokowski, RP; Pant, PV; Dadd, T; Fereday, A; Hinds, DA; Jarman, C; Filsell, W; Ginger, RS; Green, MR; Van der Ouderaa, FJ; Cox, DR. A genomewide association study of skin pigmentation in a South Asian population. *Am J Hum Genet*. 81(6):1119-1132. 2007.

Sturm, RA. Molecular genetics of human pigmentation diversity. *Hum Mol Genet*. 18(R1):R9-R17. 2009.

Sturm, RA & Frudakis, TN. Eye colour: portals into pigmentation genes and ancestry. *Trends in Genetics*. 20 (8): 327-332. 2004.

Sturm, RA & Larsson, M. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res*. 22(5):544-562. 2009.

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

Sulem, P; Gudbjartsson, DF; Stacey, SN; Helgason, A; Rafnar, T; Magnusson, KP; Manolescu, A; Karason, A; Palsson, A; Thorleifsson, G; Jakobsdottir, M; Steinberg, S; Pálsson, S; Jonasson, F; Sigurgeirsson, B; Thorisdottir, K; Ragnarsson, R; Benediktsdottir, KR; Aben, KK; Kiemeney, LA; Olafsson, JH; Gulcher, J; Kong, A; Thorsteinsdottir, U; Stefansson, K. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet.* 39 (12): 1443-1452. 2007.

Sulem, P; Gudbjartsson, DF; Stacey, SN; Helgason, A; Rafnar, T; Jakobsdottir, M; Steinberg, S; Gudjonsson, SA; Palsson, A; Thorleifsson, G; Pálsson, S; Sigurgeirsson, B; Thorisdottir, K; Ragnarsson, R; Benediktsdottir, KR; Aben, KK; Vermeulen, SH; Goldstein, AM; Tucker, MA; Kiemeney, LA; Olafsson, JH; Gulcher, J; Kong, A; Thorsteinsdottir U, Stefansson K. Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet.* 40(7):835-837. 2008.

SWGMAT (ScientificWorking Group on Materials Analysis). Forensic human hair examination guidelines. *Forens Sci Communicat. 7(2).* Available at http://www.fbi.gov/hq/lab/fsc/backissu/april2005/standards/2005_04_standards02.htm. 2005.

Takamoto, T; Schwartz, B; Cantor, LB; Hoop, JS; Steffens, T. Measurement of iris color using computerized image analysis. *Curr Eye Res.* 22, 412–419. 2001.

TASI. Technical Advisory Service for Images. *Colour theory: Understanding and modelling colour.* University of Bristol. Available at http://www.tasi.ac.uk/advice/creating/pdf/colour. pdf, accessed December 5, 2007. 2004.

Terman, JS & Terman, M. Photopic and scotopic light detection in patients with seasonal affective disorder and control subjects. Biol Psychiatry. 46: 1642–1648. 1999.

Theos, AC; Truschel, ST; Raposo, G; Marks, MS. The Silver locus product Pmel17/gp100/Silv/ME20: controversial in name and in function. *Pigment Cell Res.* 18(5): 322–336. 2005.

Toyofuku, K; Valencia, JC; Kushimoto, T; Costin, GE; Virador, VM; Vieira, WD; Ferrans, VJ; Hearing, VJ. The etiology of oculocutaneous albinism (OCA) type II: the pink protein modulates the processing and transport of tyrosinase. *Pigment Cell Res.* 5: 217–224. 2002.

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

Valverde, P; Healy, E; Jackson, I; Rees, JL; Thody, AJ. Variants of the melanocyte stimulating hormone receptor gene are associated with red hair and fair skin. *Nat Genet.* 11 (3): 328–330. 1995.

Vaughn, MR; Brooks, E; van Oorschot, RA; Baindur-Hudson, S. A comparison of macroscopic and microscopic hair color measurements and a quantification of the relationship between hair color and thickness. *Microsc Microanal.* 15(3):189-193. 2009a.

Vaughn, M; van Oorschot, R; Baindur-Hudson, S. A comparison of hair colour measurement by digital image analysis with reflective spectrophotometry. *Forensic Sci Int* 183, 97–101. 2009b.

Vaughn, M; van Oorschot, R; Baindur-Hudson, S. Hair color measurement and variation. *Am J Phys Anthropol.* 137(1): 91–96. 2008.

Voight, BF; Kudaravalli, S; Wen, X; Pritchard, JK. A map of recent positive selection in the human genome. *PLoS Biol*. 4: e72. 2006.

Watabe, H; Valencia, JC; Yasumoto, K; Kushimoto, T; Ando, H; Muller, J; Vieira, WD; Mizoguchi, M; Appella, E; Hearing, VJ. Regulation of tyrosinase processing and trafficking by organellar pH and by proteasome activity. *J Biol Chem*. 279(9):7971-7981. 2004.

Werrett, DJ. The new frontiers of forensics. *Policy Today*. 11: 29–31. 2005.

Weskamp, G; Mendelson, K; Swendeman, S; Le Gall, S; Ma, Y; Lyman, S; Hinoki, A; Eguchi, S; Guaiquil, V; Horiuchi, K; Blobel, CP. Pathological neovascularization is reduced by inactivation of ADAM17 in endothelial cells but not in pericytes. *Circ Res*. 106(5):932-940. 2010.

Wielgus, AR & Sarna, T. Melanin in human irides of different color and age of donors. *Pigment Cell Res*. 18, 454–464. 2005.

Willems, SH; Tape, CJ; Stanley, PL; Taylor, NA; Mills, IG; Neal, DE; McCafferty, J; Murphy, G. Thiol isomerases negatively regulate the cellular shedding activity of ADAM17. *Biochem J*. 428(3):439-450. 2010.

Williamson, SH; Hubisz, MJ; Clark, AG; Payseur, BA; Bustamante, CD; Nielsen, R. Localizing recent adaptive evolution in the human genome. *PLoS Genet*. 3:e90. 2007.

Yamamoto, T; Suganami, T; Kiso-Narita, M; Scherle, PA; Kamei, Y; Isobe, M; Higashiyama, S; Ogawa, Y. Insulin-induced ectodomain shedding of heparin-binding epidermal growth factor-like growth factor in adipocytes in vitro. *Obesity*. 18(10): 1888-1894. 2010.

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

7.3. APÊNDICE III

(Coluna “Considerações sobre genética da pigmentação humana e a política de cotas raciais” no Portal Educação – disponível em http://www.portaleducacao.com.br/direito/artigos/20931)

CONSIDERAÇÕES SOBRE GENÉTICA DA PIGMENTAÇÃO HUMANA E A POLÍTICA DE COTAS RACIAIS

02/11/2012 23:12:00

Recentemente tem se discutido em alguns veículos de comunicação sobre a criação de cotas raciais para o serviço público. Para o ingresso em muitas universidades brasileiras esta política já se faz presente. Entretanto, gostaria de aproveitar este espaço para fazer algumas considerações a respeito do tema. Para começar, digo que o fato de ser "branco" ou "negro", a princípio, não possui qualquer correlação com a capacidade de ser aprovado num vestibular ou num concurso público. Todos possuem a mesma capacidade. Entretanto, o fato de ser pobre ou não faz uma grande diferença.

Nas cotas raciais, qual o ponto de corte ou critérios utilizados? A auto-declaração? O sistema de auto-declaração é muito falho para a implementação desta política, por motivos óbvios. Talvez, num futuro próximo, seja possível estimar a cor da pele de um indivíduo somente através da análise do DNA. Mas é bom lembrar que uma estimativa não é uma certeza. Além disso, os fatores genéticos que conferem a cor da pele estão sujeitos a grande influência ambiental. Por exemplo, existem pessoas que são mais fáceis de bronzear do que outras, tendo o mesmo tom de pele. Isto se deve à ancestralidade individual e também, possivelmente, a diferenças sutis em outros fatores que influenciam o tom de pele, como a proporção de gordura epitelial, distribuição dos vasos sanguíneos na pele, etc. Além disso, um possível teste de DNA para cor de pele somente seria interessante para definir os extremos do espectro dos tons de pele: branco ou negro. O espectro da cor de pele é muito grande! A cor "morena", por exemplo, apresenta tonalidades diversas. Interpretar as informações genéticas para distinguir entre estes tons ainda parece uma atividade impossível. Ainda não sabemos como criar um critério para isso.

Sabe-se também da correlação existente entre a cor da pele e a ancestralidade. Com o estudo do DNA, é possível verificar a porcentagem aproximada de ancestralidade individual existente no material genético de cada pessoa. Entretanto, também neste contexto das ações afirmativas, como definir quem se beneficiaria com estes resultados? Penso que a maioria absoluta de brasileiros (provavelmente todos, se pesquisarmos o suficiente) pode ter alguma proporção de material genético Africano no seu DNA. Isto se considerarmos o fato de que os ancestrais dos humanos modernos vieram da África, berço da humanidade. Isto é história e já é sabido pela evolução humana. Além disso, em populações resultantes da mistura demográfica entre Asiáticos, Europeus, Africanos e Nativos americanos, como a população Brasileira o é, definir a ancestralidade individual para a finalidade das ações afirmativas pode ser bastante problemático. Por exemplo, comparando-se o DNA entre duas pessoas "morenas" escolhidas aleatoriamente, sendo uma delas com tom de pele visivelmente mais claro, é plenamente possível encontrar maior ancestralidade genética Africana nesta do que naquela com tom de pele mais escuro. Diante deste fato, nos resta a dúvida: Como não errar na implementação das cotas raciais? O que queremos realmente discutir? Critério racial ou Critério socioeconômico? Para responder estas questões, é necessário ver exatamente o foco do problema, a fim evitar o agravamento da questão.

Na perspectiva da educação brasileira, o que é alarmante é o fato de ainda existir grande divergência educacional no ensino do país. O que se observa são muitas universidades públicas com qualidade de um lado, em comparação com um ensino público básico (ou fundamental) e secundário (ou médio) ainda bastante deficientes. Desta forma, muitos estudantes que vieram do ensino básico e secundário das escolas públicas parecem não cumprir o nível de conhecimento exigido para o ingresso no nível superior de qualidade e acabam se deparando com uma dificuldade enorme para passar (quando conseguem) num vestibular de universidade (pública ou particular) de excelência, e acabam, muitas vezes, não cursando o ensino superior que gostaria, devido à sua base educacional limitada. Em outra dimensão estão os estudantes que tiveram um ensino básico e secundário de excelência, e que, muitas vezes, conseguem a

maioria das vagas nas melhores universidades e/ou concursos existentes.

Nesta perspectiva, podemos observar que a solução do problema é a melhoria do ensino básico e secundário no país. Para isto, o Poder público deve implementar com urgência políticas de investimento na educação brasileira, por exemplo: incrementar os salários dos professores, melhorar a infraestrutura das escolas e universidades, fiscalizar com afinco o ensino público e privado no país, criar escolas e professores "intinerantes", que possam alcançar alunos da zona rural e da população abaixo da linha de pobreza, bem como injetar mais recursos no nível superior para sua expansão e contínuo desenvolvimento. Estas opções são mais trabalhosas. Mas, sem isso, o Estado Federativo só tem a perder com a criação somente de medidas paliativas.

Mas, já que a política é criar cotas ao invés de melhorar a educação no país, que seja discutida a criação de cotas exclusivamente para estudantes provenientes de escolas públicas, se for o caso, sejam eles brancos ou negros. Embora, na atual realidade social do Brasil, 70% dos pobres são negros, também existem brancos pobres. Não podemos nos esquecer disto! A lei de cotas raciais acaba excluindo o homem branco e pobre também como sujeito passivo da exclusão educacional existente no país. Isso é inadmissível e grave, e deveria ser discutido largamente na sociedade antes da sua implantação. A política de resguardar os direitos de uma classe não pode dar margem e/ou esquecimento aos direitos do outro grupo. Pelo menos, não deveria. Os Estados Unidos, um dos primeiros países a adotar a política de cotas, estão reavaliando sua política de ingresso nas universidades públicas baseada em cotas raciais. Lá, alguns estados já estão modificando suas políticas em direção às cotas socioeconômicas! Este critério parece ser mais sensato.

A igualdade dos povos e o respeito à mulher, ao negro, ao homossexual, enfim, o respeito ao ser humano como um todo deve ser divulgado e disseminado, a fim de que não se repita eventos históricos frutos de mentalidades retrógradas e de políticas negativas. Entretanto, na divulgação e disseminação desse respeito a grupos específicos, temos que tomar o cuidado de não criar ainda mais políticas excludentes e discriminatórias, a fim de que realmente todos possam ter educação, saúde, lazer, qualidade de vida e igualdade de condições para conquistar um nível de vida digno.

Leituras Adicionais:
http://www.jornaldaciencia.org.br/Detalhe.jsp?id=84659
http://www.jornaldaciencia.org.br/Detalhe.jsp?id=84660
Autor: Caio Cesar Silva de Cerqueira

7.4. APÊNDICE IV

(Entrevista “A cor da pele escrita no DNA”, concedida ao repórter Salvador Nogueira da Revista Pesquisa FAPESP, Março de 2012. Disponível em http://revistapesquisa.fapesp.br/2012/03/29/a-cor-da-pele-escrita-no-dna/).

A cor da pele escrita no DNA

Testes permitem descobrir características raciais sem a presença das pessoas

SALVADOR NOGUEIRA | Edição 193 - Março de 2012

© NELSON PROVAZI

É um estudo que contesta o que se costuma ouvir nas aulas de genética no ensino médio: "Características como cor da pele dependem de relações complexas entre muitos genes, o que praticamente inviabiliza identificar a aparência (fenótipo) de um indivíduo a partir de sua constituição genética (genótipo)". Embora a primeira premissa esteja correta, pesquisadores brasileiros mostram que é possível sim determinar a pigmentação da pele com base nos genes. Por ora, com modestos 60% de acerto.

Esse resultado, obtido pelo grupo da geneticista Maria Cátira Bortolini, da Universidade Federal do Rio Grande do Sul (UFRGS), não teria sido possível se não estivéssemos no início da era da genômica personalizada, em que se tornou viável economicamente sequenciar o conjunto completo de genes de indivíduos e deixá-los disponíveis em banco de dados na internet. Com essa disponibilidade, a equipe brasileira não precisou sequenciar o DNA de ninguém e fez o trabalho com dados genômicos encontrados em bases públicas mundo afora.

A equipe da UFRGS usou no total informações de 30 genomas individuais. Alguns eram de pessoas conhecidas, como os geneticistas norte--americanos Craig Venter e James Watson, o que permitiu confrontar os dados genéticos com as características fenotípicas (aparência). Outros eram de indivíduos anônimos, cujos fenótipos os pesquisadores estimaram a partir das características físicas das etnias a que as pessoas pertenciam. No estudo também foram analisados os genomas de um paleoesquimó e de quatro hominídeos arcaicos: três neandertais; e um hominídeo de Denisova – na realidade, uma mulher que viveu na Sibéria 40 mil anos atrás e que pode pertencer a uma espécie desconhecida do gênero *Homo*.

Para garimpar os dados úteis à pesquisa nessa imensa sopa de letrinhas – cada genoma é formado por dois conjuntos de 23 cromossomos com um total de 3 bilhões de pares de bases nitrogenadas A, T, C e G –, os pesquisadores usaram uma metodologia desenvolvida pelo geneticista Caio Cesar Silva de Cerqueira, primeiro autor do estudo, aceito para publicação no *American Journal of Human Biology*. "Esse trabalho é um desdobramento de meu projeto de doutorado que diz respeito a genes de coloração em populações humanas", conta Cerqueira, que é orientado por Cátira.

E não se deve subestimar o tamanho da tarefa. "Uma das maiores dificuldades foi encontrar um modo de analisar tamanha quantidade de dados ao mesmo tempo", diz Cerqueira. "Segundo nosso conhecimento, não existe ainda aparato estatístico que faça isso de maneira simplificada."

Por essa razão, a primeira missão da equipe foi reduzir a análise aos trechos de DNA

que pudessem dar maior confiabilidade às estimativas. O grupo trabalhou basicamente com as diferenças genéticas conhecidas como polimorfismos de nucleotídeo único – *single nucleotide polymorphisms* ou SNP. Os SNPs representam diferenças genéticas em que apenas uma letra da sequência foi trocada. "Tivemos que filtrar os dados para trabalhar só com os mais palpáveis e concretos", explica Cerqueira.

Seleção de dados

O ponto de partida foi identificar 346 SNPs distribuídos em 67 genes, pedaços de genes desativados (pseudogenes) e regiões intergênicas (nem todos os segmentos de DNA constituem genes, alguns só ocupam espaço na sequência, com função ainda não esclarecida). Todos esses SNPs estavam em regiões do genoma associadas à pigmentação de cabelos, olhos e pele. O passo seguinte foi ver quais desses SNPs já tinham seu efeito genético descrito na literatura. Dos 346, sobraram 124.

Ainda assim, havia um problema: o genoma se compõe de duas cópias de cada gene, uma do pai e outra da mãe. Quando as versões do gene são diferentes entre si, predizer o efeito que a combinação terá no organismo é bem complicado. Por isso, os pesquisadores se concentraram nos SNPs cujos alelos (versões diferentes encontradas simultaneamente no organismo) estivessem presentes nas duas cópias do mesmo trecho de cada genoma. "Perdemos uma boa quantidade de informações fazendo isso, mas optamos por essa abordagem mais conservadora", conta Cerqueira.

Uma coisa é certa. A metodologia usada é excelente para determinar se o indivíduo tem ou não sardas, acúmulo de pigmento comum em loiros e ruivos. A taxa de acertos na previsão para os 11 genomas cujo fenótipo era bem conhecido (a identidade do proprietário era sabida) foi de impressionantes 91%.

© NELSON PROVAZI

Contudo, conforme as sutilezas aumentaram, o nível de acerto diminuiu. O método previu corretamente em 64% dos casos o tom de pele, dividido em duas categorias: claro e escuro. A taxa de predição foi de 44% para a cor do cabelo (preto, castanho, ruivo e loiro), 36% para a cor dos olhos (preto, castanho, verde e azul). Quando todas as características foram levadas em consideração, a média de acerto ficou em 59%.

Os pesquisadores também calibraram o nível de acerto incluindo 19 genomas de indivíduos cuja etnia permitia estimar o provável fenótipo. Com uma base de 30 genomas, os índices de acerto mudaram um pouco. Diminuíram ligeiramente para sardas (83%), pele (60%) e cabelos (42%). Mas aumentaram para os olhos (67%), elevando a média final para 63%.

A primeira impressão que o estudo deixa é de que não dá mais para classificar como impossível prever traços físicos com base na análise do DNA. E a segunda é que ainda falta avançar bastante para que o nível de precisão melhore a ponto de o teste se tornar útil.

Trata-se de uma tecnologia que pode revolucionar, por exemplo, a ciência forense. Imagine se, a partir de uma amostra de DNA encontrada numa cena de crime, a polícia pudesse criar um retrato detalhado de um suspeito. Ainda estamos longe desse estágio tecnológico, mas, para Cátira, já terminou a fase do "e se pudéssemos?" e estamos chegando à etapa do "como faremos?".

"O grande desafio é entender como funciona a interação entre os vários genes e seus alelos, bem como de seus produtos, as proteínas", afirma Cátira. "Em outras palavras, o quanto o efeito de um alelo que se encontra em determinado ponto da rota é alterado pela presença de outra variante em outro gene da rede de pigmentação. Os estudos dessas conexões estão só começando a surgir e não fazemos ideia de como tudo está conectado, resultando em determinado fenótipo", diz a pesquisadora da UFRGS.

Como complicação adicional, ainda é preciso levar em conta os efeitos epigenéticos – a influência de fatores ambientais sobre os padrões de expressão de certos genes sem alterar o DNA em si. "Os desafios permanecem grandes", comenta Cátira. "Mas, como o conhecimento científico cresce exponencialmente, tenho esperança de que avanços importantes ocorram nos próximos anos."

Retratos da evolução

Enquanto a tecnologia não chega ao ponto de ajudar no trabalho policial, os pesquisadores já começam a usá-la para tentar compreender melhor como se deu a evolução do gênero *Homo*. Afinal de contas, estudos como esse ajudam a verificar o quanto a diferença de pigmentação entre os grupos humanos é resultado de pressão exercida pela seleção natural ou consiste em variações surgidas ao acaso, neutras do ponto de vista evolutivo.

Em trabalho anterior, ligado a outra característica, o grupo de Cátira havia mostrado que um gene associado à configuração dos membros em seres humanos se mantém exatamente igual em mais de uma centena de amostras de DNA, vindas de pessoas de diversas partes do globo. Esse gene acumulou 16 alterações desde que os seres humanos e os chimpanzés se separaram na árvore evolutiva e permaneceu idêntico nos neandertais – *Homo neanderthalensis*, espécie aparentada do *Homo sapiens*, com quem conviveu até cerca de 30 mil anos atrás, antes de desaparecer. A conclusão é que esse gene é extremamente importante e por isso foi conservado igual por tanto tempo, dada a pressão evolutiva que existe sobre ele.

Agora os pesquisadores também podem fazer análises semelhantes com relação à pigmentação da pele, dos olhos e dos cabelos e ver que papel evolutivo os genes relacionados a essas características podem ter tido. Antes mesmo de qualquer análise do DNA, muitos pesquisadores já pensavam que deve ter havido grande pressão evolutiva para que os humanos vivendo sob o intenso sol africano tivessem mais melanina na pele – e mais proteção contra a radiação solar nociva –, enquanto quem vivia no norte da Europa dificilmente precisasse de grandes quantidades desse pigmento para escapar dos danos causados pela exposição ao sol. Estudos como o do grupo de Cátira ajudam a compreender o que moldou essa e outras adaptações.

Um dos resultados surpreendentes do novo estudo foi mostrar que entre os nean-dertais possivelmente já havia diferenças na cor da pele e dos cabelos. A análise das variantes genéticas dos neandertais – foram sobrepostos trechos do genoma de três fêmeas para obter um genoma completo – sugere que uma era ruiva e duas tinham cabelo castanho e pele mais escura. Todas tinham olhos castanhos.

Esse resultado contrasta com o de um trabalho anterior, conduzido por Carles Lalueza-Fox, da Universidade Pompeu Fabra, em Barcelona. Em um estudo

publicado em 2007 na *Science*, o grupo espanhol mostrou que o material genético de dois neandertais – um encontrado na Espanha e outro na Itália – apresentavam alterações genéticas similares às que determinam pele clara e cabelos ruivos em seres humanos. "Temos conhecimento de poucos genomas ou de porções do genoma desses hominídeos e, mesmo assim, essa variação aparece", diz Cátira.

Se estiver correta, a análise da equipe da UFRGS indica que entre os neandertais a pigmentação poderia variar de modo semelhante ao que ocorre com os seres humanos. "Isso seria bem razoável e indicaria que essa característica pode ser típica do gênero *Homo* e não da espécie humana", comenta Cátira. Ela própria, porém, adverte que é preciso ter cautela na interpretação dos dados. "Não podemos descartar problemas metodológicos, como contaminação com DNA humano e troca de bases *post-mortem*, no sequenciamento de genomas de espécies extintas", completa.

Essa observação toca num ponto importante: há limitações na análise do material genético de fósseis. Por exemplo, talvez jamais seja possível investigar o DNA dos primeiros seres humanos que colonizaram o que hoje é o Brasil e que teriam vindo da Ásia pelo estreito de Bering entre 20 mil e 12 mil anos atrás. "O problema é que o clima daqui não permite preservar DNA da mesma forma que no frio da Europa", explica Fabricio Rodrigues dos Santos, biólogo da Universidade Federal de Minas Gerais. "Se por muita sorte encontrarem algum esqueleto preservado em algum lugar muito especial na América do Sul com mais de 8 mil anos e houver DNA, pode ser possível sim prever alguns fenótipos."

E quanto a Luzia – o fóssil humano encontrado nos anos 1970 pela arqueóloga francesa Annette Laming-Emperaire em Lagoa Santa, Minas Gerais – que hoje detém o recorde de mais antigo da América, com estimados 11.400 a 16.400 anos de idade? "No caso de Luzia, diria que é impossível, pois tentaram várias vezes, enviaram aos Estados Unidos e à Europa, mas nunca conseguiram gerar nenhuma sequência de DNA", diz Santos. Por mais que a genética tenha o poder de iluminar o passado humano, algumas lacunas inevitavelmente permanecerão. Pelo menos até a próxima grande revolução científica.

Artigo científico

CERQUEIRA, C.C.S. *et al*. Predicting *Homo* pigmentation phenotype through genomic data: From Neanderthal to James Watson. **American Journal of Human Biology**. No prelo.

7.5. APÊNDICE V

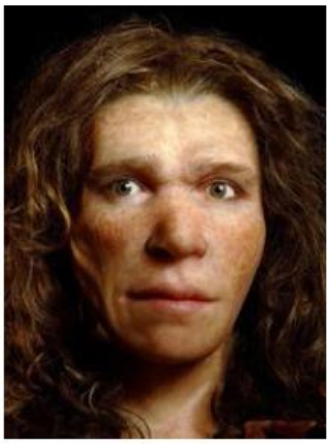
(Entrevista “Were Some Neandertals Brown-Eyed Girls?”, concedida à repórter Traci Watson da Revista *ScienceNOW*, Março de 2012. Disponível em http://news.sciencemag.org/sciencenow/2012/03/were-some-neandertals-brown-eyed.html).

Were Some Neandertals Brown-Eyed Girls?

by Traci Watson on 19 March 2012, 3:29 PM | 0 Comments

In museums around the world, reproductions of Neandertals sport striking blue or green eyes, pale skin, and gingery hair. Now new DNA analysis suggests that two of the most closely studied Neandertals—a pair of females from Croatia—were actually brown-eyed girls, with brunette tresses and tawny skin to match. The results could help shed new light on the evolution of the family that includes both modern humans and Neandertals, who died out some 30,000 years ago.

ENLARGE IMAGE



True visage? Not all Neandertals sported this look, according to new research.

Credit: Philippe Plailly & Atelier Daynes / Photo Researchers, Inc

The study has provoked deep skepticism among several outside researchers, however, who criticize numerous aspects of its methodology. The results also run contrary to other genetic evidence and to a long-held hypothesis that Neandertals, who lived mostly in northern latitudes, must've had light skin to get enough vitamin D.

But even scientists who have doubts about the new research say it still provides food for thought. "Neandertals occupied a wide geographical range," says John Hawks of the University of Wisconsin, Madison, who was not involved in the study and who is also studying the physical traits of ancient humans, so "it's likely that they were variable in pigmentation. ... We are really at the first step."

The new study, to be published in the *American Journal of Human Biology* later this spring, looks at the genomes of three female Neandertals from Croatia. Their DNA was the basis of the first effort to compile a complete Neandertal genetic sequence, which was published in 2010.

The researchers focused their attention on 40 well-studied stretches of genetic material that help determine pigmentation in living people. A particular form of the gene known as *TPCN2*, for example, bestows brown hair in modern humans; any other form means hair that's another color.

One complication is that traits such as hair color are controlled by multiple genes. To determine the cumulative impact of multiple genes on one trait, the authors assumed they could simply add together the impact of individual genes. The female Neandertal known as Vi33.26, for example, had seven genes for brown eyes, one for "not-brown" eyes, three for blue eyes, and four for "not-blue eyes." By the researchers' reckoning, that means a six-gene balance in favor of brown and a negative balance for blue, so Vi33.26's eyes were probably brown. According to this method, all three Neandertals had a dark complexion and brown eyes, and although one was red-haired, two sported brown locks.

Scie

Live Chat: The Genes We Eat Thursday 3 p.m. EDT

Study author Tábita Hünemeier of Brazil's Universidade Federal do Rio Grande do Sul says she's not surprised by the results. "There was a large population of Neandertals in Europe," she says. "It's impossible that an entire population has red hair or blue eyes."

She and her colleagues validated their technique, in part by applying to it the genomes of 11 modern humans whose photos and DNA are publicly available. Nearly 60% of the formula's predictions matched the subjects' actual physical appearance, the authors say. The team considers that accuracy rate satisfactory, given the complexity of the genetics behind skin color and other physical traits.

But experts caution against giving those museum exhibits a makeover just yet. The problem with the additive technique is that different genes have different levels of impact, says Carles Lalueza-Fox of Spain's Institute of Evolutionary Biology in Barcelona. In 2007, he authored a paper showing that two Neandertals, one from Italy and one from Spain, carried a genetic variation likely to bestow pale skin and red hair. He argues that some pigmentation genes have such a powerful effect that they override the combined contributions of many weaker genes—a phenomenon that would render the new study's simple gene addition inaccurate. The lighter skin color seen in Europeans, for instance, is due almost entirely to a single gene, he says. "We know that there are some genes that have a very strong effect" on physical appearance, he says.

Another problem, Hawks says, is that the study focuses on the effects of genetic variations found in modern humans. But Neandertals' hair and skin tones were almost certainly influenced by genetic variations unique to Neandertals, who were a species different from modern humans. So the study doesn't, and can't, consider many of the factors that would've influenced how Neandertals looked.

Hünemeier responds that her team looked for new genetic variations unique to Neandertals and other ancient humans and came up empty-handed. She also says that other recent work confirms that it's possible to compute the impact of large numbers of genes using simple arithmetic.

Although Hünemeier and her critics differ on the methods her team used, they agree that the stereotypical view of Neandertals is too narrow. Lalueza-Fox says Neandertals probably had brown eyes and a variety of hair colors, and Hawks thinks Neandertals living in places such as Israel may have had darker skin than their European counterparts.

The uncertainty may not last much longer. Hünemeier and her critics alike think the growing trove of information about the DNA of ancient humans will soon reveal Neandertals' true colors. New genetic information is being generated on "hundreds of individual paleopopulations," Hünemeier says. "In 5 years we will have an ocean of information to study."

*Follow **Science**NOW on Facebook and Twitter*

Posted in Anatomy, Morphology, Biomechanics | Genetics

7.6. APÊNDICE VI

(Trabalhos apresentados em congressos relacionados com a presente Tese)

Influência do gene *TPCN2* no fenótipo da pigmentação da pele e cabelos em amostra do Rio Grande do Sul.

Santos, Franciele Lausch dos[1]; Cerqueira, Caio Cesar Silva de[2]; Ramallo, Virgínia[2]; Hünemeier, Tábita[2]; Bortolini, Maria Cátira[3].

[1]Bolsista BIC/ UFRGS no Laboratório de Evolução Humana e Molecular e estudante de Graduação em Ciências Biológicas da Universidade Federal do Rio Grande do Sul.
[2]Pós-Graduandos no Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul.
[3]Orientadora no Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande Sul.
Palavras-chave: Associação, Genótipos, Pigmentação Humana.

Agências Financiadoras: CAPES, CNPq, FAPERGS.

Identificar as variantes genéticas relacionadas com pigmentação humana tem sido o objeto de interesse em estudos evolutivos, médicos e forenses. A pigmentação de cabelos, pele e olhos apresenta um amplo espectro de variação fenotípica em algumas populações e esses fenótipos são explicados principalmente pela variação na produção do pigmento orgânico melanina. Esse pigmento é sintetizado nas células denominadas melanócitos, dentro de organelas denominadas melanossomos. Há dois tipos de melanina: eumelanina (espectro marrom-preto) e feomelanina (espectro amarelo-vermelho), que são sintetizadas de acordo com o ambiente químico e hormonal presente nos melanossomos. Diferentes genes têm sido associados à variação fenotípica da pigmentação em algumas populações, incluindo o gene *TPCN2* (*Two Pore segment Channel 2*). Este gene possui 25 éxons, está localizado no cromossomo 11q, e codifica uma proteína com 752 aminoácidos. A função desta proteína é principalmente influenciar no transporte de cálcio e controlar os níveis de pH intracelular, o que, por sua vez, interfere na produção e balanço eumelanina/feomelanina. O objetivo do presente estudo é avaliar a influência de dois polimorfismos (rs3750965 – K376R e rs3829241 – G734E) no gene *TPCN2* sobre a pigmentação da pele e dos cabelos em uma amostra do Rio Grande do Sul. Neste trabalho foram genotipados 234 voluntários provenientes do projeto CANDELA (Consórcio para Análise da Diversidade e Evolução Latino-Americana). Estes voluntários foram qualificados para cor de cabelo em 4 categorias: Ruivo, Loiro, Loiro escuro/Castanho claro ou Castanho escuro/Preto. Com relação à cor de pele, foi realizada uma medida indireta da quantidade de melanina na pele destes indivíduos através de um espectrofotômetro de reflectância. Este aparelho mensura o Índice de Melanina (IM), que vai de 20 (pele clara) a 100 (pele escura) para pele humana. Este projeto foi aprovado no Comitê de Ética em Pesquisa do HCPA e da UFRGS. As frequências alélicas e genotípicas foram estimadas por contagem, assim como a verificação do Equilíbrio de Hardy-Weinberg (EHW) e as análises estatísticas foram realizadas pelo software SPSS 17.0. Para o cálculo de desequilíbrio de ligação, foi utilizado o programa MLocus. As frequências alélicas de ambos os polimorfismos analisados estavam de acordo com o esperado para EHW. Na análise de associação dos genótipos com cor de cabelos, nenhum dos polimorfismos analisados apresentou associação significativa (rs3829241 – χ^2=10,855, p=0,059; rs3750965 – χ^2=7,346, p=0,259), embora observamos que no polimorfismo rs3829241, a associação esteve próxima da significância. Da mesma maneira, na análise dos genótipos com cor de pele, nenhum dos polimorfismos se mostrou significante (rs3750965 – F=2,393, p=0,094; rs3829241 – F=1,207, p=0,301). Embora tenha sido observado que ambos os polimorfismos estão ligados (D' = 0,972 e r^2=0,241, p<0,001) na análise do MLocus, não realizamos análises adicionais, uma vez que nenhum dos polimorfismos foram associados isoladamente com as características. Os resultados do presente trabalho mostram ausência de associação dos polimorfismos rs3750965 e rs3829241 no gene *TPCN2* nos fenótipos de cor de pele e cor de cabelos em uma amostra da população do Rio Grande do Sul.

Análise de associação de três polimorfismos no gene *MC1R* com cor de cabelo em uma amostra de gaúchos.

Krause, Carla Daiana Demkio Volasko[1]; Cerqueira, Caio Cesar Silva de[2]; Ramallo, Virgínia[2]; Hünemeier, Tábita[2]; Bortolini, Maria Cátira[3].

[1]Bolsista voluntária de Iniciação Científica no Laboratório de Evolução Humana e Molecular e estudante de Graduação em Ciências Biológicas da Universidade Federal do Rio Grande do Sul.
[2]Pós-Graduandos no Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul.
[3]Orientadora no Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande Sul.

Palavras-chave: Genética Forense, Genótipo, Fenótipo.

Agências Financiadoras: CAPES, CNPq, FAPERGS.

A predição de características externas visíveis com a utilização de marcadores moleculares tem sido muito estudada pelos pesquisadores e promete se tornar uma ferramenta valiosa para o uso na genética forense. Assim, pode-se tentar predizer o fenótipo de um suspeito a partir de amostras de DNA encontradas na cena do crime. Para que isso seja possível, é necessário buscar marcadores associados com o fenótipo em populações específicas através de estudos de associação genética. Existem alguns estudos relatando a associação de SNPs (*single nucleotide polymorphisms*) de vários genes à cor da pele, dos cabelos e dos olhos em populações européias e asiáticas, mas pouquíssimos estudos em populações com ancestralidade miscigenada, como a população brasileira. Um desses genes é o *MC1R*, o qual codifica uma proteína transmembrana denominada receptor de melanocortina 1, expressa na superfície dos melanócitos. Esse gene possui apenas 1 éxon, mas desempenha um papel importante na pigmentação normal da pele e dos cabelos, controlando qual o tipo de melanina que é produzida por essas células. O *MC1R* é um gene muito polimórfico, apresentando mais de 70 variantes em indivíduos de ascendência européia. Algumas dessas variantes (rs1805007 - R151C, rs1805008 - R160W e rs1805009 - D294E) estão associadas a cabelo ruivo, pele clara e/ou presença de sardas. O objetivo do presente estudo é verificar a existência de associação das variantes supracitadas no *MC1R* com a cor dos cabelos em uma amostra de gaúchos. As amostras foram coletadas de acordo com protocolo do projeto CANDELA (Consórcio para Análise da Diversidade e Evolução Latino-Americana). Os voluntários da pesquisa assinaram o Termo de Consentimento Livre e Esclarecido antes da coleta de material biológico, como requisito para participação no estudo. Foram genotipadas 235 amostras provenientes de voluntários do estado do Rio Grande do Sul. Os indivíduos foram classificados qualitativamente em 4 categorias de cor de cabelos (Ruivo, Loiro, Loiro escuro/Castanho claro ou Castanho escuro/Preto), conforme literatura corrente. As frequências alélicas e genotípicas foram estimadas por contagem, bem como o Equilíbrio de Hardy-Weinberg (EHW). As análises de qui-quadrado foram feitas no programa SPSS, versão 17.0. Foi também realizada análise de Desequilíbrio de ligação para os três polimorfismos utilizando o programa Mlocus. As frequências genotípicas dos três polimorfismos estudados no presente trabalho estavam de acordo com o esperado para o EHW. Na análise de associação, o polimorfismo rs1805009 foi associado com cor de cabelo (χ^2=8,998; p=0,032). Os outros dois polimorfismos não estão associados com cor de cabelo na presente amostra (rs1805007 – χ^2=3,666, p=0,278; rs1805008 – χ^2= 5,427, p=0,128). Na análise de desequilíbrio de ligação, nenhum dos três polimorfismos no gene *MC1R* se apresentou ligados (rs1805007/rs1805008 - p=0,471; rs1805007/rs1805009 - p=0,677; rs1805008/rs1805009 - p=0,628). Desta forma, o presente estudo indicou associação do marcador genético rs1805009 com cor de cabelo na população gaúcha, e ausência de associação dos marcadores rs1805007 e rs1805008 com esta característica.

É possível predizer fenótipo de cor de pele em uma população miscigenada através da análise de genótipos?

Krause, Carla Daiana Demkio Volasko[1]; Cerqueira, Caio Cesar Silva de[1]; Ramallo, Virgínia[2]; Salzano, Francisco, M[1]; Hünemeier, Tábita[1]; Bortolini, Maria Cátira[1].

[1] Laboratório de Evolução Humana e Molecular do Departamento de Genética e do Programa de Pós-Graduação em Genética e Biologia Molecular -UFRGS.

E-mail: c_aioc@yahoo.com.br

Palavras-chave: Associação, Genética Forense, Genótipo-Fenótipo.

Agências Financiadoras: CAPES, CNPq, FAPERGS.

Vários estudos sobre a relação genótipo-fenótipo dos marcadores genéticos para pigmentação humana têm sido realizados especificamente em populações Européias e Asiáticas, com a finalidade de tentar predizer fenótipo de pigmentação através das análises de DNA para possível aplicação em Genética Forense. Entretanto, pouco se sabe a respeito da relação destes marcadores genéticos em populações com diferentes níveis de miscigenação, tal como a população Brasileira. O objetivo do presente trabalho é estudar a associação de 5 SNPs em dois genes *OCA2* e *TYR* (rs1800407, rs1800401 e rs1800414; rs1126809 e 1042602, respectivamente), buscando estabelecer a relação dos alelos destes com a cor de pele de 559 voluntários do projeto CANDELA (*Consórcio para Análise da Diversidade e Evolução Latino-Americana*) provenientes de diferentes regiões do país, predominantemente dos estados do Rio Grande do Sul e Bahia. O índice de melanina (IM), medida indireta da cor da pele estimada através de um espectrofotômetro de reflectância, varia de 20 (pele mais clara) a 100 (pele mais escura), em humanos. Este projeto foi aprovado no Comitê de Ética em Pesquisa do Hospital de Clínicas de Porto Alegre (HCPA) e da UFRGS. As frequências alélicas e genotípicas foram estimadas por contagem, assim como a verificação do Equilíbrio de Hardy-Weinberg (EHW) e as análises estatísticas foram realizadas pelo software SPSS 17.0. As frequências genotípicas dos 5 polimorfismos pesquisados estavam de acordo com o esperado para o EHW. Na análise de associação, fizemos três análises separadas, para fins comparativos, que correspondem a perfis amostrais com diferentes níveis de história demográfica: análise da amostra total (n=559), análise da amostra proveniente da população do Rio Grande do Sul – RS (n=347), que são predominantemente de ascendência européia, e análise da amostra da população das outras regiões do país (n=212), predominantemente de indivíduos da Bahia - BA. Quando utilizamos os 559 indivíduos numa mesma análise, 4 polimorfismos foram estatisticamente significativos, exceto o rs1800414 - *OCA2* ($F=0,017$; $p=0,895$). Quando analisamos somente os indivíduos provenientes do RS, somente o polimorfismo rs1042602 (*TYR*) foi estatisticamente significativo ($F=4,517$; $p=0,012$). Quando analisamos os indivíduos das outras regiões do país, predominantemente indivíduos do estado da Bahia, somente os polimorfismo rs1042602 – *TYR*, rs1126809 – *TYR* e rs1800407 – *OCA2* foram estatisticamente significantes ($F=8,512$, $p<0,001$; $F=5,765$, $p=0,004$; $F=6,712$, $p=0,010$, respectivamente). É notável que nas análises realizadas, observamos para maioria dos SNPs com associação significante, que os genótipos heterozigotos possuíram um IM intermediário, sugerindo um efeito aditivo dos marcadores. Este fato também é observado por outros estudos na literatura, analisando amostras de populações nativas. Análises adicionais serão realizadas para elucidar o motivo dessas diferentes associações nos três grupos amostrais, visto que

diferentes fatores de confundimento, tal como estruturação populacional, podem estar associados aos resultados aqui relatados.

Análise de marcadores genéticos e predição de características físicas: o caso do *Homo neanderthalensis* e do espécime de Denisova

Cerqueira, CCS[1*]; Paixão-Côrtes, V[1]; Hünemeier, T[1]; Salzano, FM[1]; Bortolini, MC[1].

[1] Instituto de Biociências/ Departamento de Genética. Programa de Pós-Graduação em Genética e Biologia Molecular – PPGBM. Universidade Federal do Rio Grande do Sul – UFRGS.

** E-mail: c_aioc@yahoo.com.br*

Palavras-chave: Marcadores moleculares, Neandertal, Pigmentação, Retrato Molecular, SNPs.

Variação na pigmentação da pele, cabelos e olhos figuram entre os exemplos mais visíveis da diversidade fenotípica humana. Recentemente, a genética forense deu um passo a mais no uso da tecnologia do DNA - pesquisadores têm estudado marcadores moleculares para tentar predizer o fenótipo de possíveis suspeitos. Em outras palavras, procura-se realizar um retrato molecular do indivíduo a partir da amostras de DNA encontrado em cenas de crime. Na literatura, encontramos vários estudos científicos associando cor de pele, cabelos e olhos a variações de base única (SNPs) em genes candidatos. A partir destes estudos, e tendo em vista as publicações em 2010 e 2011 das seqüências preliminares dos genomas completos do *Homo neanderthalensis* (Neandertal) e de outro, um espécime de hominídeo sem classificação taxonômica clara, encontrado na Caverna de Denisona (Rússia), tentamos ir mais longe. Portanto, o objetivo do presente estudo é estabelecer o provável fenótipo da cor da pele e cabelo dos hominídeos cujo genoma foi publicado recentemente. Para isso, analisamos 20 SNPs localizados em 10 genes que atuam em diferentes estágios do processo de produção de pigmentos, incluindo melanogênese, a estabilização e transporte de enzimas durante a síntese de melanina, bem como no balanço entre a síntese de diferentes tipos de melanina. Os marcadores selecionados foram: MATP/SLC45A2 (rs16891982, rs26722, rs6867641), TYR (rs1042602, rs1126809), ASIP (rs6058017), SLC24A5/ NCKX5 (rs1426654), ADAM17 (rs1524668), OCA2 (rs1800414, rs1800407, rs1800401), EDAR (rs3827760), HERC2 (rs1129038), MC1R (rs1805007, rs1805008, rs1805009, rs4785763), TPCN2 (rs3829241, rs3750965, rs35264875). Estes marcadores foram então procurados nas seqüências destes hominídeos nos UCSC Genome Browser Gateway (http://genome.ucsc.edu/cgi-bin/hgGateway e http://genome.ucsc.edu/Neandertal/). Para 19 marcadores, o alelo ancestral (também presente no genoma do chimpanzé; *Pan troglodytes* - UCSC Genome Browser on Chimp Oct. 2010 - CGSC 2.1.3/panTro3 - Assembly) foi encontrado em ambos hominídeos. Curiosamente, no SNP rs1805009 (Asp294His - G>C) do gene MC1R, no genoma do Neandertal, um alelo diferente está presente (A). Porém, a troca de aminoácido continua a mesma (Asp294His). Este SNP foi associado anteriormente a pele clara e cabelos ruivos em populações européias. Nosso estudo com estes 20 marcadores identifica pelo menos dois possíveis fenótipos diferentes para o Neandertal e para o homem de Denisova. O Neandertal seria possivelmente ruivo com pele clara, enquanto o homem de Denisova teria cabelos castanhos ou pretos. Com a utilização de mais marcadores poderá ser possível fazer um retrato molecular e fenotípico mais preciso dos membros extintos do gênero *Homo*.

Instituições financiadoras: CNPq, CAPES, FAPERGS.

Genetic markers for skin color forensic DNA phenotyping in the Brazilian admixed population

Cerqueira, CCS[1]; Ramallo, V[1]; Hünemeier, T[1]; Barbosa, AAL[2]; Schüler-Faccini, L[1,3]; Salzano, FM[1]; Bortolini, MC[1].

[1] Programa de Pós-graduação em Genética e Biologia Molecular (PPGBM), Laboratório de Evolução Humana e Molecular (LEHM) da Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre-RS.
[2] Departamento de Ciências Biológicas, Universidade Estadual do Sudoeste da Bahia (UESB), Jequié-BA.
[3] Instituto Nacional de Genética Médica Populacional (INAGEMP), Porto Alegre-RS.

c_aioc@yahoo.com.br

Key-words: Admixed populations, Forensic practice, Human skin color, Population differences, Set os SNPs

The human pigmentation pathway can be explained by a complex network controlled by many genes, as well as by environmental, mechanical, and epigenetic factors; thus, to understand it is a challenging task. Although dozens of genes have previously been associated with human skin color, our knowledge about this trait is still incomplete. Particularly limited is the number of studies with populations off the Europe-North American axis, and seldom until now admixed populations were considered. The present study was planned to help fill this gap. The objective was to verify the possible association of 18 SNPs located within ten genes/pseudogene regions (*ADAM17*rs1524668, *AFG3L1*rs4785763, *ASIP*rs6058017, *HERC2*rs1129038, *MC1R*rs1805007, *MC1R*rs1805008, *MC1R*rs1805009, *OCA2*rs1800401, *OCA2*rs1800407, *OCA2*rs1800414, *SLC24A5*rs1426654, *SLC45A2*rs6867641, *SLC45A2*rs26722, *SLC45A2*rs16891982, *TPCN2*rs3750965, *TPCN2*rs3829241, *TYR*rs1042602, and *TYR*rs1126809), with the MI (Melanin Index) in two admixed Brazilian populations (*Gaucho*, n=354; *Baiano*, n=149) with different histories of geographic and ethnic colonization. Some of these polymorphisms have never been studied in the context of human skin color normal variation. DNA extractions were done using the salting out method from total blood, while genotyping procedures were performed by Taqman™. MI measurements were taken on the proximal medial portion of both arms to generate a single average value (range in humans from 20, associated to the fairest skin, to 100, to darkest skin). The SPSS software, version 17.0, was used to calculate ANOVA or the non-parametric Kruskal-Wallis test, comparing MI versus genotypes. We identified 4 markers associated with MI in the *Gaucho* sample (*HERC2*rs1129038, *MC1R*rs1805009, *SLC24A5*rs1426654, and *SLC45A2*rs16891982), and 6 in the *Baiano* (*HERC2*rs1129038, *OCA2*rs1800407, *SLC24A5*rs1426654, *SLC45A2*rs16891982, *TYR*rs1042602, and *TYR*rs1126809), but only 3 were significantly associated in the two samples (*HERC2*rs1129038, *SLC24A5*rs1426654, *SLC45A2*rs16891982). Therefore, only these 3 should be preliminarily chosen as being of forensic significance, since they consistently show the association independently of the population considered. Furthermore, to consider possible population differences is also important to choose an appropriate set of SNPs as phenotype predictors in forensic practice.

Financial Support: CAPES, CNPq, FAPERGS, The Leverhulme Trust, and The Biotechnology and Biological Sciences Research Council.

7.7. APÊNDICE VII

(Artigo submetido para *PLOS Genetics* com dados gerais do Projeto CANDELA)

GENETIC ANCESTRY, PHENOTYPIC DIVERSITY AND SELF-PERCEPTION OF ANCESTRY IN 6,938 LATIN AMERICANS

Andrés Ruiz-Linares[1], Kaustubh Adhikari[1], Victor Acuña-Alonzo[1,2],Mirsha Quinto[3], Claudia Jaramillo[4], William Arias[4], Macarena Fuentes[5], María Pizarro[5], Paola Everardo[2,6], Francisco de Avila[2,6], Jorge Gómez-Valdés[6], Paola León-Mimila[6], Tábita Hunemeier[7], Virginia Ramallo[7], Caio C. Silva de Cerqueira[7], Mari-Wyn Burley[1], Marcelo Zagonel de Oliveira[7], Mauricio Roberto Veronez[8], Marta Rubio-Codina[9], Orazio Attanasio[9,10], Sahra Gibbon[11], Carla Gallo[12], Giovanni Poletti[12], Javier Rosique[13], Lavinia Schuler-Faccini[7], Francisco M. Salzano[7], Maria-Cátira Bortolini[7], Samuel Canizales-Quinteros[6,14], Francisco Rothhammer[5], Gabriel Bedoya[4], David Balding[1] and Rolando Gonzalez-José[3]

[1] Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, UK
[2] National Institute of Anthropology and History, Mexico City, México
[3] Centro Nacional Patagónico, CONICET, Puerto Madryn, Argentina
[4] Universidad de Antioquia, Medellín, Colombia
[5] Instituto de Alta Investigación Universidad de Tarapacá, Programa de Genética Humana ICBM Facultad de Medicina Universidad de Chile and Centro de Investigaciones del Hombre en el Desierto, Arica, Chile
[6] Facultad de Medicina and Facultad de Química, UNAM, Mexico City, México
[7] Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil
[8] Remote Sensing and Digital Imaging Laboratory, Graduate Program on Geology, Vale do Rio dos Sinos University, São Leopoldo, Brazil
[9] The Institute for Fiscal Studies, London, UK.
[10] Department of Economics, University College London, UK.
[11] Department of Anthropology, University College London, London, UK.
[12] Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Perú
[13] Departamento de Antropología. Facultad de Ciencias Sociales y Humanas. Universidad de Antioquia, Medellín, Colombia
[14] Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán", Mexico City, México

Corresponding author:
Andrés Ruiz Linares
Department of Genetics, Evolution and Environment
University College London
Darwin Building
Gower Street
London WC1E 6BT
United Kingdom
Tel: +44 (0)20 3108 4003 Fax: +44 (0)20 7679 7193
e-mail: a.ruizlin@ucl.ac.uk

ABSTRACT

We report a characterization of the genetic ancestry (African, European and Native American), phenotypic diversity and self-perception of ancestry in 6,938 Latin American individuals recruited in Brazil, Chile, Colombia, México and Perú. Phenotypes examined include pigmentation (of hair, eyes and skin), hair type (curliness) and greyness, male pattern baldness, and a range of anthropometric and facial features. We observe extensive variation in ancestry between individuals and geographic regions. We find significant effects of ancestry (P< 10-3) on most phenotypes examined, consistent with the geographic structure of these traits in Africa, America and Europe. Significant association was detected for two of the markers typed: rs260690 (in the Ectodysplasin A Receptor, *EDAR* gene) is associated with hair type (P-value 2 10^{-16}) and rs717225 (in the Rho guanine nucleotide exchange factor 1, *ARHGEF1* gene) with skin pigmentation (P- value 1.58 10^{-12}). We observe a significant correlation between self-perception and genetic estimates of ancestry and find evidence for significant effects of pigmentation and social factors in biasing self-perceived relative to genetically estimated ancestry.

The history of the countries in the region collectively known as "Latin America" has involved the coming together of individuals with diverse geographic origins of which the demographically most prominent encounter has been that occurring between the American natives and people arriving mostly from the Iberian peninsula and West (Sub-Saharan) Africa starting in 1492(1-3). In the past several years there has been an increasingly detailed genetic characterization of the resulting history of genetic admixture in Latin America(4-9). Amongst the most prominent conclusions from these studies is the observation of an extensive variation in genetic ancestry between Latin American individuals and regions, and that admixture involved predominantly immigrant men and native women. These findings generally agree with information on pre-Columbian native population density across the Americas and the historical patterns of overseas immigration to specific geographic regions of the continent(2). The sex-bias in admixture is also consistent with documentary evidence showing that it was mostly men that migrated to the New World during the colonial period, particularly in the first century of settlement, when admixture was likely to have been most extensive(1, 10).

Studies relating the genetic diversity of Latin Americans to phenotypic variation have so far been scant(11-13). Such genotype-phenotype correlation analyses are of considerable biological importance as they would contribute to elucidating the genetic basis of variation in human physical appearance, particularly of traits differing in frequency between the major populations involved in Latin American admixture. The identification of genes influencing these traits is of wide-ranging relevance. Firstly, human physical attributes are used extensively in evolutionary analyses, ranging from human origins (i.e. "humans" are defined as "anatomically modern") to studies on potential environmental adaptations, with the implicit assumption that such attributes reflect a specific genetic make-up. However, the poorly defined genetic basis of these traits limits many of the inferences made in such

analyses. Secondly, identifying genes for variation in human physical appearance is of biomedical relevance as such genes would constitute candidates potentially involved in pathological variation in those traits. It has also been shown that certain physical features are associated with a variable disease risk, but a biological explanation for these associations is often unclear. For instance, pigmentation has been associated with risk of Parkinson's disease(14) and skin cancer(15), male pattern baldness with heart disease(16) and prostate cancer(17), and facial features with general health history(18, 19). Thirdly, identification of genetic variants influencing human physical features promises to be useful in the prediction of an individual's appearance from genetic data, an application of considerable relevance in forensics(20), although one which is not without challenges and problematic social implications. Similarly, characterizing the genetic basis of variation in human physical appearance, particularly that occurring between individuals with different continental origins, has cultural ramifications outside of forensics and biomedicine. Variation in physical appearance has long been used in attempts to define "human races" and historically many of these traits have been highly charged socially(21). Evaluating factors influencing individual perception of variation in these traits, their association with popular notions of ancestry and identity and how these are informed or transformed by the genetic ancestry information, is thus of considerable relevance to the social sciences and for society at large(22-24).

Here we examine the correlation between genetically estimated African/European/Native American ancestry, variation in individual physical appearance and self-perception of ancestry in 6,938 individuals ascertained in Brazil, Chile, Colombia, Mexico and Perú. We also evaluated the possibility of a direct effect of the genetic markers used for ancestry estimation in variation in physical appearance and found novel associations of skin pigmentation with *ARHGEF1* and of macroscopic hair type with *EDAR*.

MATERIALS AND METHODS

Study subjects

Recruitment took place mainly in five cities: Mexico City (México), Medellín (Colombia), Lima (Perú), Arica (Chile) and Porto Alegre (State of Rio Grande do Sul, Brazil). With the exception of Chile, most subjects recruited in these cities were students and staff from the Universities participating in this research. In Chile about 2/3 of the subjects recruited were professional soldiers from a battalion in Arica. In Brazil some samples were also collected in smaller towns of Rio Grande do Sul, as well as in towns of the states of Bahia and Rondonia . Adult subjects of both sexes were invited to participate mainly through public lectures and media presentations. We obtained ethics approval from: Escuela Nacional de Antropología e Historia (México), Universidad de Antioquia (Colombia), Universidad Peruana Cayetano Heredia (Perú), Universidad de Tarapacá (Chile), Universidad Federal do Rio Grande do Sul (Brazil) and University College London (UK). Each participant provided written informed consent. Blood samples were collected by a certified phlebotomist and DNA extracted following standard laboratory procedures.

Phenotypic data

A physical examination of each volunteer was carried out by the local research team using the same protocol and instruments at all recruitment sites. We recorded: height, weight, head, hip and waist circumference, cheilion-cheilion width and sellion-gnation height. All measurements were obtained in duplicate and the mean of the two measurements retained for further analyses. Using visual inspection we recorded eye colour into five categories (1-blue/grey, 2-honey, 3-green, 4-light brown, 5-dark brown/black), and natural hair colour into four categories (1-red/reddish, 2-blond, 3-dark blond/light brown or 4-brown/black)(25). Balding in males was recorded using a modified Hamilton scale(26-29) as: 0) no hair loss, 1) frontal baldness only, 2) frontal hair loss with mild vertex baldness, 3) frontal hair loss with

moderate vertex baldness, and 4) frontal hair loss with severe vertex baldness. Similarly, greying was recorded along a five point scale: 0) for no greying, 1) for predominant non-greying, 2) for ~50% greying, 3) for predominant greying and 5) for totally white hair(29, 30). Macroscopic hair type was categorized by visual inspection as straight, wavy, curly or frizzy (31). A quantitative measure of constitutive skin pigmentation (the Melanin Index, MI) was obtained using the DermaSpectrometer DSMII reflectometer (Cortex Technology, Hadsund, Denmark)(32, 33). Measurements were obtained on both inner arms and the mean of the two readings used in the analyses.

Five digital photographs of the face: left side (-90°), left angle (-45°), frontal (0°), right angle (45°), right side (90°) were taken from ~1.5 meters at eye level using a Nikon D90 camera fitted with a Nikkor 50mm fixed focal length lens. The photographs were annotated manually with 36 anatomical landmarks and 3D landmark coordinates extracted using the software Photomodeler (http://www.photomodeler.com/ Eos Systems Inc, Vancouver, Canada) (Supplementary Figure S1). Landmark configurations were superimposed by Generalized Procrustes Analysis(34) and the nuisance effects of BMI, sex, and age removed by multiple regression of the shape coordinates. Principal Components (PCs) of the residual landmark coordinates were obtained using the software MorphoJ(34). To ease visualization of the 3D shape changes associated with principal components, we generated a file with spatial information for the i-th PC score produced by MorphoJ and obtained deformation surfaces via thin a plate spline algorithm using the Landmark software (35).

Socioeconomic position and self-perception of ancestry

A structured questionnaire was applied to each volunteer so as to obtain information on socioeconomic position (SEP) and on self-perception of ancestry. We obtained two indicators of SEP. The first is maximum educational attainment (EA), categorized as:

none/primary (1), technical (2), secondary (3), university (4) and post-graduate (5). The second is a wealth index (WI) obtained from a list of items taken as indicators of standard of living. These items were: home ownership, number of bathrooms at the place of residence, ownership of cars, bicycles, fridge/freezer/dishwasher, TVs, radios, CD/DVD players, vacuum cleaner, washing machine and availability of domestic service. We used polychoric principal component analysis to examine the variability of the sample and retained the first principal component as a WI.

Volunteers were asked for their self-perception of African, European, and Native American ancestry. This was understood as a rough estimation of the proportion of their ancestors that could have a particular continental origin. We proposed a five point scale, expressed as 20% per cent brackets (and in words): 1) 0-20% (none or very low), 2) 20-40% (low), 3) 40-60% (moderate), 4) 60-80% (high) and 5) 80-100% (very high or total). Allowance was provided for recording ancestries outside Africa, Europe and America. The questionnaire also recorded information on: the place of birth of the volunteer and (when known) of their ancestors (up to grandparents), various racial/ethnicity categories (such as those used colloquially and by the census in each country), as well as on individual's attitudes towards human genetics research, ancestry studies and racial discrimination. Analyses of the data from the full socioeconomic questionnaire will be published separately.

Genetic admixture estimation

In order to select 30 markers highly informative for estimating African/European/Native American ancestry, we started from the list of 5,000 markers, highly informative for world-wide continental ancestry estimation, identified by Paschou et al (2010)(36) using the approach of Rosenberg et al. (2003)(37) based on the worldwide CEPH-HGDP cell panel genotyped with Illumina's 610 chip (including data for about 600,000

SNPs(38). The full list of these 5,000 markers is at: http://www.cs.rpi.edu/~drinep/HGDPAIMS/WORLD_5000_INFAIMs.txt. Of these, allele genotype data is available in Native Americans for 3,848 markers(39), of which 2,392 markers have been placed on subsequent Illumina chip products. This subset of markers was retained for selection of those to be typed here in order to facilitate subsequent data comparison and integration. We ranked the 2,392 markers based on allele frequency differences in European-Native American or European-African samples. Amongst markers with the highest inter-continental allele-frequency differences we selected those with lowest heterozygosity in Native Americans (so as to reduce the effect of variable allele frequencies between Native Americans on ancestry estimation). Of the final set of 30 markers retained, 13 are monomorphic in 408 Native Americans (from 47 populations Mexico Southwards), the rest have minor allele frequencies ranging from 0.01 to 0.15 (median 0.06) in that group of populations. This set of 30 markers produces individual ancestry estimates in a sample of Colombians recently included in a genome-wide association study that used Illumina's 610 chip(40) with correlations of ~0.7 compared with estimates obtained from the chip data, and identical mean population estimates. The list of markers typed is provided in Supplementary Table S1. Genotyping was carried out by KBiosciences (Hoddesdon, Herts, UK; http://www.kbioscience.co.uk/). Proportions of African, European and Native American ancestry (and their standard errors) were estimated using the ADMIXTURE software(41), using supervised runs (unsupervised runs produced very similar estimates) and standard errors obtained by bootstrap using the program default parameters. Data from a total of 876 individuals sampled in putative parental populations were used in ancestry estimation. These were selected from HAPMAP, the CEPH-HDGP cell panel(38) and from published Native American data(39) as follows: 169 Africans (from 5 populations from Sub-Saharan West Africa), 299 Europeans (from 7 West and South European populations) and 408 Native

Americans (from 47 populations from Mexico Southwards). The full list of the putative parental population samples (and their sizes) is provided in Supplementary Table S2.

Geographic analyses

Information on birthplaces for volunteers, their parents and grandparents was consolidated into a unique list of locations (for all countries except Peru, due to a large number of missing data in that country). These placenames were organized into three fields: city/municipality, region/state and country. Data consistency was checked by comparing field names, e.g. name of a state into the list of city/states for each country. The list of placenames was converted via the Google Maps Geocoding API (https://developers.google.com/maps/documentation/geocoding/) into geographic coordinates. Query results were checked for consistency to verify that the geocoding process has identified all places properly. The GeodesiX software (http://www.calvert.ch/geodesix/) was used for the geocoding query. The final birthplaces datafile containing coordinates was represented into a searchable map using Google Fusion Tables (http://www.google.com/drive/apps.html#fusiontables).

Geographic maps displaying individual admixture estimates were obtained using the software ArcGis 9.3 (http://www.esri.com/software/arcgis). The cartographic database was geo-referenced to the SIRGAS geodesic system (Geocentric Reference System for the Americas http://www.ibge.gov.br/home/geociencias/geodesia/sirgasing/index.html) using a Universal Transverse Mercator projection. Krirging interpolation was used to adjust for uneven sampling across geographic regions. Corel-DRAW X3 (Corel Corporation, Ottawa, Canada) was used to edit the map images.

Statistical analyses

Linear or logistic regression was used to examine the effect of genetic ancestry or marker genotypes, and selected covariates on phenotypic variation. Analyses were performed using the R package (www.r-project.org)(42).

RESULTS

Genetic ancestry

Documentary information indicates that the geographic origin of the individuals contributing to the current gene pool of Latin America varied greatly. It is also well established that there is genetic substructure in the regions where these individuals originated. Here we use the terms ''African(s)'', ''European(s)'' and "Native American(s)" to refer broadly to different geographic origins without assuming a biological homogeneity, or strict distinctiveness, of these regions. Our main aim here is to obtain ancestry estimates consistently across samples as an heuristic allowing us to examine patterns of variation in ancestry and their correlation with variation in phenotype and self-perception. We did not attempt to perform a refined examination of the recent geographic origin for the ancestors of individuals. Since our focus is not on fine-grained individual ancestry but on the correlation of broad continental ancestry with other variables we estimated it with a relatively small number of (highly informative) markers in a large number of individuals. We also do not attempt to draw a comprehensive picture of the admixture landscape of Latin America (the individuals studied here are not meant to be representative, in a statistical sense, of the countries where the samples were collected). In what follows we use the country names simply as labels indicating where recruitment took place.

The sample collected in Brazil has a highly predominant European ancestry (median 87%) and relatively low Native American and African ancestries (both medians <10%) (Table 1). The samples collected in Chile, Mexico and Perú have highest median Native

American ancestry (46%, 58% and 66% respectively) and lowest median African ancestry (1%, 3% and 1% respectively). The sample collected in Colombia is the most diverse in showing a predominantly European ancestry (median 62%) but substantial Native American admixture (median 28%) and the highest African component (median 7%). There is considerable heterogeneity across ascertainment sites in the distribution of individual ancestry estimates (Figure 1). The Mexican sample is characterized by the widest spread of individual ancestry along the Native American-European axis. By contrast the Chilean sample, shows a greater concentration around the median. The Brazilian, and particularly the Colombian samples, show a greater spread of ancestry along the African axis, with some individuals having a high (>70%) estimated African ancestry. Although most individuals studied here were born near the main ascertainment sites, there is some geographic spread of their birthplaces. When examining the geographic distribution of individual ancestry estimates within each country we observe considerable regional variation but with some broad trends being apparent (Figure 2). In Brazil (Figure 2A) European ancestry is highest in the south, African ancestry in the North East and Native American ancestry in the west. In Chile (Figure 2B) highest European ancestry is seen around the main urban areas of the north and centre of the country with Native ancestry predominating elsewhere and a north to south gradient of decreasing African ancestry. In Colombia (Figure 2C), African ancestry is highest in the coastal regions (particularly on the Pacific), Native ancestry being highest in the South West and in the East of the country, European ancestry being highest in central areas. In Mexico (Figure 2D) Native American ancestry is particularly high in the south, the north showing the highest proportion of European Ancestry (Figure 2b). Other than geographic variation, individual genetic ancestry also shows significant variation with SEP, European ancestry increasing with both higher educational attainment and wealth. This trend is

significant in the global sample (r= 0.32 P-value <2.2 10^{-16} and r=0.06 P-value = 3.05 10^{-05} for EA and WI respectively) and at each of the sampling sites (Supplementary Note S1).

Phenotypic diversity and genetic ancestry

Table 2 and Figure 3 summarize the distribution of quantitative and categorical phenotypes (respectively) across samples. As is the case for the ancestry estimates, we find a significant heterogeneity for most of the phenotypes examined. For traits differing in frequency between continental populations, we usually observe a variation in frequency across samples consistent with the ancestry estimates (Table1). For instance, the (highly predominantly European) Brazilian sample shows lighter pigmentation (of skin, hair and eyes) and greater height than the samples from the other countries, which show higher Native American ancestry. When comparing eye color groups we observe an increase in European ancestry for the sequence of eye colors: black, brown, green, honey and blue (Figure 4). Amongst hair color groups (Figure 4) there is an increase in European ancestry from black to brown to blond, the red hair group showing a European ancestry between the black and brown hair color groups. Similarly, there is an increase in European ancestry for the groups with straight, wavy and curly hair, with the frizzy hair group showing the lowest European ancestry (Figure 4). Relative to the overall sample the frizzy hair group has markedly higher African ancestry (median of 0.42 v. 0.04).

Regression of phenotypic variation on genetic ancestry demonstrates a significant effect (P-value < 10-3, using a Bonferroni multiple testing correction) for most of the traits examined (Table 3 and Figure 5). Accounting for sex, sampling site, age, and SEP, higher European ancestry is associated with increased height, lower body mass index (BMI) and central obesity (ICO), lighter pigmentation (or hair, skin and eyes), greater hair curliness and male pattern baldness (with hair greying approaching significance). Similarly, higher African ancestry is associated with increased height, greater skin pigmentation and greater hair

curliness type. The proportion of phenotypic variance explained by ancestry is highest for skin pigmentation (19%) and is also substantial for hair shape (10%) and color of eyes and hair (4% and 5 %, respectively) but considerably smaller for the other phenotypes. We also observe highly significant effects of EA (P-value 3.87 10^{-13}) and age (P-value <2 10^{-16}) on height with height increasing for younger individuals at a rate ~1 cm for every 10 years age difference (Supplementary Note S2). Accounting for height and BMI (as well as for the previous covariates) ancestry also has significant effects on facial variation both in terms of size and shape. Higher European ancestry is associated with an overall smaller face size (as estimated by centroid size). We also examined the effect of ancestry on all pair-wise inter-landmark distances (427, after averaging the 203 distances between symmetric landmarks on both sides of the face), which combine shape and size information. Of these, a total of 133 and 2 show significant effects of European and Africa ancestry, respectively (P-values < 10-6, assuming a conservative Bonferroni multiple testing correction; Supplementary Table S3). The most significant effects of European ancestry (P<-10) mainly involve distances between landmarks placed on the lips and nose. Face shape variation, independent of size, was assessed via Principal Component Analysis of procrustes coordinates. Significant effects of European ancestry were detected for PCs 1 and 3-5, while African ancestry impacts on PCs 1, 2 and 4 (Table 3, Supplementary Note S3 and Supplementary Figure S2). These 5 PCs account for ~55% of the variation in face shape captured by the 36 landmarks placed on the facial photographs, with ancestry explaining up to 5% of the variance in PC scores (for PC4). Examination of the correlation between inter-landmark distances and facial PCs (results not shown), indicates that the highest correlation of distances between landmarks of the lips and nose is seen with PC4, consistent with this PC showing the largest proportion of variance explained by ancestry (Table 3). Regression of individual shape scores (providing a single value for shape change) obtained from landmark coordinates as proposed by Drake and

Klingenberg(43), agree with a significant effect of ancestry on overall face shape (Table 3 and Figure 3).

Since the markers used for ancestry estimation have allele frequencies that are highly differentiated between Africans, Europeans and Native Americans, they constitute candidate variants potentially impacting on traits with variable frequencies between these populations. We therefore tested for association between the 30 markers typed and the phenotypes showing ancestry effects (with genome-wide ancestry, age, sampling site, SEP and sex as covariates). Marker rs260690, located in the first intron of the EDAR gene, was found to have a highly significant effect on hair type ($p= 10^{-16}$). Examination of HapMap data for Mexican Americans (MEX) indicates high linkage disequilibrium (D'=0. 98) between this SNP and rs3827760 an A to G variant coding for a p.Val370Ala substitution in EDAR. The derived G allele at rs3827760 has a high frequency in CHB+JPT (0.95) but is rare in CEU+TSI (frequency 0.05) and this allele has recently been associated with increased cross-sectional hair thickness in East Asians as well as in mouse models(44-46). Marker rs356652, 8 Mb from marker rs260690, also shows significant association with hair type (P-value = $1.94\ 10^{-07}$). These two marker are in moderate linkage disequilibrium (D'=0.52 in the full sample) and accounting for rs260690 in the regression model markedly attenuates the evidence for association of hair type with rs356652 (P-value 0.004), confirming that this signal of association stems from LD with variants in the EDAR gene. We also found a highly significant association (P-value = $1.58\ 10^{-12}$) between marker rs717225, located in the third intron of the ARHGEF1 gene, and skin pigmentation. In the HAPMAP dataset, the ancestral C allele at rs717225 is found at a frequency of 90% in YRI and 0.01 in CEU and is absent from available Native American data (Supplementary Table S1).

Genetic ancestry and self-perception

We observe a moderate, highly significant, correlation of self-perception of continental ancestry with the genetic estimates (ranked into five bands at 20% increments): America: r= 0.48 P <2.2 10^{-6}, Europe: r=0.48 P <2.2 10^{-6}, Africa: r=0.32 P <2.2 10^{-6}. However, there is a trend at increasing Native American and African ancestry for self-perception to exceed the genetic estimates (Figure 6). Correspondingly, at decreasing European ancestry there is a trend for the genetic estimates to exceed self-perception (Figure 6). Regression analysis of the difference between individual self-perceived and genetically estimated ancestry (Table 4) indicates a significant effect of skin pigmentation on this difference: individuals with lower skin pigmentation tend to overestimate their European ancestry while individuals with higher pigmentation overestimate their Native American and African ancestries. Similarly, lighter eye color is associated with a significant overestimation of European ancestry and underestimation of Native American ancestry (but not African ancestry). Overall, men overestimate their European ancestry and underestimate their African ancestry, relative to women. There is also a significant variation between countries in that individuals from the Chilean, Colombian, and Mexican samples tend to overestimate their Native American ancestry and underestimate their European and African ancestries relative to the Brazilian sample.

DISCUSSION

The paleo-anthropological record shows that the physical appearance of humans has undergone considerable changes throughout evolution, presumably reflecting the interplay of various environmental and genetic factors. Previous studies have also shown that current human phenotypic diversity shows some clinal geographically structure, usually to a similar degree as seen with "neutral" genetic markers. For instance, it has been estimated that about 90% of both craniofacial variation and genetic diversity is present within continents and only

about 10% of this variation is accounted for by differentiation between continents(47). For certain phenotypes however (such as skin pigmentation), there often is greater differentiation between than within populations and this pattern has been interpreted as consistent with the action of selection on these phenotypes. In the case of skin pigmentation possibly due to variable levels of sunlight exposure, which could have a range of physiological effects potentially affecting survival(48, 49). Consistent with this scenario, recent genome-wide analyses aimed at detecting selection at the molecular level have often implicated genes involved in pigment synthesis(12, 50, 51).

Recently admixed populations, such as Latin Americans, offer an opportunity to characterize the genetic basis of traits showing geographic structure between the populations that contributed to admixture in the region; in this case, mostly Native Americans and individuals arriving from Africa and Europe. The analyses performed here illustrate the great diversity of Latin Americans, with an extensive variation in genetic ancestry and physical appearance. Although not designed to be statistically representative of the populations sampled, the geographic distribution of the ancestry estimates obtained here are consistent with the demographic history of the regions examined. The relatively low global estimates of African ancestry are consistent with the sampling sites not being located in areas of major historical African immigration. The relatively high Native American ancestry seen in the Peruvian, Mexican and Chilean samples, is consistent with the high pre-Columbian Native population density in Meso-America and the Andes, while the Colombian and Brazilian samples come from regions with comparatively lower pre-Columbian Native population densities(2, 3). In addition, the Brazilian sample is predominantly from a region (Rio Grande do Sul) that has received a large influx of European immigrants in the last two centuries(2, 3, 52). Although the sampling sites examined are not located in areas of major African immigration, the samples from Colombia and Brazil show the largest estimates of African

ancestry, consistent with the greater African immigration to those two countries, relative to Perú, México and Chile and these concentrate particularly in the coastal regions of Colombia and the Brazilian North East(2, 53-55). Within each country the geographic distribution of individual ancestry estimates is also consistent with regional demographic history. In Mexico most of the current Native population concentrates to the South of the country, European immigration having concentrated mostly in the centre and North of the country. In Colombia Native populations occupy predominantly peripheral areas, the central, highland areas having received most European immigration, while Africans predominantly settled in coastal areas and major inland river valleys(55, 56). In Brazil, most African settlement occurred in the North East, the South having received a major European population influx in the last two centuries, while most Native populations currently concentrate to the West(2, 3, 52).

The pattern of variation we observe between phenotypes and estimated genetic ancestry is generally consistent with available information regarding the variation in frequency of these phenotypes in Natives from America, Europe and Africa, and the generally complex genetic basis of these traits. Europeans usually have lighter skin pigmentation than Native Americans or Africans and also show frequent variation in eye and hair color, which is rare elsewhere. A number of genes influencing these traits have been recently identified in European(25, 57-60). Consistent with this geographic pattern, increased European ancestry is associated with lighter skin, hair and eye pigmentation and higher African ancestry with greater skin pigmentation (Table 3 and Figure3). An unexpected pattern was found for red hair. This trait is essentially absent in Native Americans and sub-Saharan Africans, and found at non-negligible frequencies in Europeans where it is associated with very low skin pigmentation, poor tanning response and high sun sensitivity. However, the Latin American individuals with red hair examined here were found to have lower European ancestry than individuals with blond or brown hair (Figure 3), showing

ancestry values closer to the population median. This observation could relate to red hair being determined by variants at relatively few genes (most prominently the melanocortin 1 receptor MC1R)(61), thus making this trait a particularly unreliable indicator of individual genome-wide ancestry. Of the SNPs examined here, rs717225, located in the third intron of *ARHGEF1* showed a strong association with skin pigmentation. The C allele of rs717225 is seen at high frequency in Africans and at very low frequency in Europeans and has not been reported in Native Americans (Supplementary Table S1) and this allele is associated with increased pigmentation in the Latin American sample examined here. Rho GTPases play a key role in cellular processes mediated by G-protein coupled receptors and recent genome-wide RNA-mediated interference experiments have implicated *ARHGEF1* in melanogenesis(62). These experiments have shown a direct impact of this gene on the expression of tyrosinase and of *MITF* (the Microphthalmia-associated transcription factor, a master regulator of melanocyte development)(62). *MC1R* is a G-protein coupled receptor whose association with red hair (as mentioned above) is well established(58) but this is the first time that variants in *ARHGEF1* are associated with human pigmentation variation.

Macroscopic hair type (curliness v. straightness) is a highly heritable trait showing considerable variation between human populations(57, 63). The hair of East Asians and Native Americans has been described as commonly coarse in texture, with individual fibres that are straight, thick and circular in cross-section(64). We found a strong effect of ancestry on hair type, with straight hair frequency increasing in individuals with greater Native American ancestry. This observation was further refined by the finding of a highly significant association of hair type with marker rs260290 in *EDAR*. A recent study found evidence of association in Asians of the p.370Ala substitution in EDAR (rs3827760>G) with hair cross-sectional thickness(44, 45). No genotype data is currently available for marker rs3827760 in Native Americans. However, there is strong LD across the *EDAR* region and an imputation

analysis of available high-density chip data for Native Americans(39) indicates that these populations are essentially fixed for the derived rs3827760 G allele (results not shown). Position 370 in EDAR is located in an intracellular domain required for the interaction with the EDAR-binding death domain adapter protein EDARADD in a pathway involved in formation of the hair placode(65) and experimental evidence suggests that the p.370Ala variant has higher activity than the ancestral p.370Val variant(46). This variant is seen at high frequency in East Asians but is absent from Europeans and Africans. The population differentiation and haplotypic structure of the EDAR region has been interpreted as evidence of recent strong selection on this gene(50, 66, 67). The absence of the p.370Ala variant in Europeans indicates that hair straightness has evolved more than once in human populations and consistent with this, variants in the trichohyalin gene have recently been shown to be associated with hair type in Europeans(68). The nature of the selection pressure however remains unclear although it has long been argued that hair type could represent an adaptation to variable levels of heat and UV radiation or that it could have been influenced by sexual selection(57). Further studies should clarify the role of EDAR variants in the development of the hair follicle and developmentally related structures such as sweat glands and teeth, helping to shed light on the selective forces that might have shaped diversity at this gene.

Hair greying and androgenetic/senescent alopecia are amongst the most obvious manifestations of aging in humans, yet their molecular basis is poorly understood. Both processes result from a weathering of the hair shaft and the aging of the hair follicle, leading to a decrease in melanocyte function and in hair production. Experimental evidence indicates that stem cell death and oxidative stress play a key role in this process(69, 70) and recent studies emphasize the importance of Notch2 signalling in the maintenance of melanocyte stem cells and hair pigmentation(71). Changes in the hair follicle associated with hair greying and loss, therefore represent a model system in which to study “intrinsic” aging (72) and

where to test anti-aging therapies. Various lines of evidence testify to the importance of genetic factors in hair loss and greying, the most extreme being Mendelian disorders associated with hair loss or premature aging(29, 73-76). A twin study estimated the heritability of androgenetic alopecia amongst Europeans at about 79%(29) and genome-wide studies in Europeans have now identified several loci associated with this trait(77, 78). It has been estimated that androgenetic alopecia affects ~45% of Caucasian men by the age of 49 and 70% by 79. Similarly, by 50 years of age, about 50% of Europeans have ~50% grey hair(75, 76, 79). Although no similar studies have been carried out in Amerindians, classical physical anthropology studies indicate that hair greying and androgenetic alopecia are rarer, less severe and of later onset in Amerinds than in Caucasians(80, 81). In the sample examined here we do find that greater European ancestry correlates with significantly higher rates of male balding and also see an increase of European ancestry with greater hair greying but only approaching statistical significance. It will be interesting to follow-up these results as elucidating the genetic basis of these traits could help increase our understanding of the aging of the hair follicle and the health risks associated by male pattern baldness.

The effect of genetic ancestry on height and related anthropometric phenotypes (BMI, ICO) is consistent with the polygenic basis of these traits documented by recent genome-wide association studies in Europeans and the existence of trait loci with allele frequencies varying between continental populations(82-84). Our results indicate that about 1% of the variation in height in the Latin American individuals examined results from genetic ancestry variation. Height is also under strong environmental influence(18) and, in line with this, we detect a significant effect of SEP in our sample. That the two SEP indicators examined here underestimates the amount of environmental variation is suggested by the significant effect of date of birth on height, with individuals born more recently tending to be taller than older individuals. The increase in height for individuals born more recently (~0.1 cm/year)

estimated here is similar to that estimated from large longitudinal surveys in Latin America (~1 cm per decade in the last century), an observation that has been interpreted as resulting from the historical improvement in living standards across the region(18, 85, 86).

Vertebrate craniofacial morphology shows great variation within species and diversification between species, where changes in craniofacial features closely match adaptive radiations to particular ecological niches. However, the genetic basis of this variation is poorly understood. In humans, twin and family studies have estimated a heritability of 60-90% for a range of facial traits(87) and anthropological studies have documented a geographic structure for facial features, which is of a similar degree to that seen with neutral genetic data(47, 88, 89). The characterization of single gene mutations in experimental organisms and in patients with dysmorphologies, has identified common signalling cascades during vertebrate craniofacial development(87, 90-92). Work in experimental organisms has often focused in cases where there is a main direction of shape change, for instance in the offspring of populations with some differentiated anatomical structures(93-100). A somewhat related situation applies here. We observe that ancestry impacts broadly on facial features described previously as varying between Africans, Europeans and Native Americans(101-103): mainly width and height of the face, facial flatness, position of the glabella and frontotemporal points and the relative size and position of lips (a fuller description of the main aspects of face shape variation associated with each PC is presented in the Supplementary Note S3 and Supplementary Figure S2). Two recent genome-wide association studies of face shape variation in Europeans have identified a few loci associated with aspects of face shape variation(104, 105) and both studies indicate that variants in the PAX3 gene influence human face shape. These analyses are consistent with the view that facial phenotypes are influenced by a large number of genetic variants of relatively small effect size (106, 107). The ancestry effect that we detect for face shape and

size but not for cranium size agrees with the notion of a greater developmental and evolutionary constraint on neurocranium than on facial variation, as suggested by several morphological analyses (108-113). These observations are in line with proposals that facial features of anatomically modern humans include environmental adaptations related to regionally differing selection pressures associated with the late Pleistocene dispersal of our species (114-117).

The correlation of individual ancestry with the various phenotypes discussed above and the specific effect of variants in *EDAR* and *ARHGEF1* in hair type and skin pigmentation, illustrate the informativeness of Latin Americans for characterizing the genetic basis of traits geographically structured across the continents contributing to admixture in the region. Latin America also offers a special opportunity to examine the relationship between biological diversity and certain social variables. Our analyses emphasize the continuing correlation of education and wealth with genetic ancestry in the region(118, 119). Although admixture in Latin America has taken place since the beginning of the European colonial expansion (some 25 generations ago), this has occurred within the context of strong socioeconomic stratification, with European immigrants and their descendants usually having a higher socioeconomic position than people with greater Native and African ancestry(1, 120, 121). As an indicator of associated environmental variation and wellbeing (i.e. stress, nutrition, healthcare, medical history) SEP is thus an important confounder that needs to be taken into account when evaluating ancestry effects on phenotype, including disease, in these populations, as we see here for height and have reported previously in relation to type 2 diabetes(118, 119). The effect of age on height detected (consistent with historical surveys(18, 85, 86) suggests that the two indicators of SEP used here are likely to capture only part of the environmental variation affecting this trait and it will be important to develop more specific indicators of environmental variation for each phenotype that capture relevant

factors. Our data also underline the major role played by physical appearance in self-perception of ancestry. As individual genealogical information on ancestry is usually limited to 2-3 generations and since admixture in Latin America can go back over 20 generations it is likely that physical appearance will often influence self-perception of ancestry. Based on the geographic structure of genetic and phenotypic diversity it is to be expected that self-perception and genetic ancestry correlate, as we observe. However, since certain visible traits show a higher level of geographic structure than the genome average, perception of ancestry based on these traits is biased relative to genetically estimated ancestry. This is illustrated for instance, by the different effect that we see for skin pigmentation and African self-perception or for eye/hair color and European self-perception (Table 4). Our findings are thus consistent with the notion that self-perception is "skin deep" in that it is heavily influenced by certain visible attributes and biased relative to genetic ancestry. We also find evidence that this bias is influenced by social factors possibly related to a variable perception regarding the extent of geographic structure of visible phenotypes, as suggested by the significant differences that we find between countries and sexes. These observations are in line with social phenomena such as the so-called "one drop rule" in the US, according to which an individual with some African ancestry is identified as African, thus constituting an extreme form of perception bias.

As the product of biological evolution our species shows complex patterns of phenotypic and genetic variation, including some geographic structure. However, perception of the geographic distribution of human genetic and phenotypic diversity impinges upon the sensitive subjects of race and racism. Both of these notions are social constructs associated with grossly inaccurate interpretations of human biological variation. It is our belief that a fuller understanding of this variation, and of its social correlates, will help undermine the

social prejudices upon which racism is based. We trust that this study makes a contribution in that direction.

ACKNOWLEDGMENTS

We are extremely grateful to the volunteers for their enthusiastic support of this research. This work was funded by grants from the Leverhulme Trust (F/07 134/DF to A.R-L) and BBSRC (BB/I021213/1 to A.R-L). We thank Ken Weiss and Peter Wade for comments on the manuscript and Lucía Castillo and Ana Carolina Orozco for technical assistance.

REFERENCES

1. Morner M (1967) *Race Mixture in the History of Latin America* (Little Brown & Company).
2. Sanchez-Albornoz N (1974) *The population of Latin America: a history* (University of California Press, Berkeley).
3. Salzano FM & Bortolini MC (2002) *The Evolution and Genetics of Latin American Populations* (Cambridge University Press).
4. Carvajal-Carmona LG*, et al.* (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* 67(5):1287-1295.
5. Carvalho-Silva DR, Santos FR, Rocha J, & Pena SD (2001) The phylogeography of Brazilian Y-chromosome lineages 1. *Am.J.Hum.Genet.* 68(1):281-286.
6. Alves-Silva J*, et al.* (2000) The Ancestry of Brazilian mtDNA Lineages. *Am.J.Hum.Genet.* 67(2):444-461.
7. Bedoya G*, et al.* (2006) Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc. Natl. Acad. Sci. U. S. A.* 103(19):7234-7239.
8. Wang S*, et al.* (2008) Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 4(3):e1000037.
9. Bryc K*, et al.* (2010) Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci. U. S. A.* 107 Suppl 2:8954-8961.
10. Boyd-Bowman P (1976) Patterns of Spanish Emigration to the Indies until 1600. *Hispanic American Historical Review* 66:580-604.
11. Ventura Santos R*, et al.* (2009) Color, Race, and Genomic Ancestry in Brazil: Dialogues between Anthropology and Genetics. *Curr. Anthrop.* 50(6):787-819.
12. Myles S, Somel M, Tang K, Kelso J, & Stoneking M (2007) Identifying genes underlying skin pigmentation differences among human populations. *Hum. Genet.* 120(5):613-621.
13. Klimentidis YC, Miller GF, & Shriver MD (2009) Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am. J. Phys. Anthropol.* 138(4):375-383.
14. Gao X, Simon KC, Han J, Schwarzschild MA, & Ascherio A (2009) Genetic determinants of hair color and Parkinson's disease risk. *Ann. Neurol.* 65(1):76-82.

15. Duffy DL, *et al.* (2010) Multiple pigmentation gene polymorphisms account for a substantial proportion of risk of cutaneous malignant melanoma. *J. Invest. Dermatol.* 130(2):520-528.
16. Yamada T, Hara K, Umematsu H, & Kadowaki T (2013) Male pattern baldness and its association with coronary heart disease: a meta-analysis. *BMJ open* 3(4).
17. Wright JL, Page ST, Lin DW, & Stanford JL (2010) Male pattern baldness and prostate cancer risk in a population-based case-control study. *Cancer epidemiology* 34(2):131-135.
18. McEvoy BP & Visscher PM (2009) Genetics of human height. *Economics and human biology* 7(3):294-306.
19. Friedlander AH & Scully C (2010) Diagonal ear lobe crease and atherosclerosis: a review of the medical literature and oral and maxillofacial implications. *J. Oral Maxillofac. Surg.* 68(12):3043-3050.
20. Kayser M & de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nature reviews. Genetics* 12(3):179-192.
21. Gilman SL (1985) *Difference and pathology : stereotypes of sexuality, race, and madness* (Cornell University Press, Ithaca) p 292 p.
22. Cho MK & Sankar P (2004) Forensic genetics and ethical, legal and social implications beyond the clinic. *Nat. Genet.* 36(11 Suppl):S8-12.
23. Ossorio PN (2006) About face: forensic genetic testing for race and visible traits. *J. Law. Med. Ethics* 34(2):277-292.
24. M'Charek A (2008) Silent witness, articulate collective: DNA evidence and the inference of visible traits. *Bioethics* 22(9):519-528.
25. Sulem P, *et al.* (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* 39(12):1443-1452.
26. Hamilton JB (1951) Patterned loss of hair in man; types and incidence. *Ann. N. Y. Acad. Sci.* 53(3):708-728.
27. Norwood OT (1975) Male pattern baldness: classification and incidence. *South. Med. J.* 68(11):1359-1365.
28. Lotufo PA, Chae CU, Ajani UA, Hennekens CH, & Manson JE (2000) Male pattern baldness and coronary heart disease: the Physicians' Health Study. *Arch. Intern. Med.* 160(2):165-171.
29. Rexbye H, *et al.* (2005) Hair loss among elderly men: etiology and impact on perceived age. *J. Gerontol. A. Biol. Sci. Med. Sci.* 60(8):1077-1082.
30. Morton DJ, Kritz-Silverstein D, Riley DJ, Barrett-Connor EL, & Wingard DL (2007) Premature graying, balding, and low bone mineral density in older women and men: the Rancho Bernardo study. *J. Aging Health* 19(2):275-285.
31. Comas J (1984) *Manual de Antropologìa Fisica* (Universidad Nacional Autonoma de Mèxico, Mèxico).
32. Clarys P, Alewaeters K, Lambrecht R, & Barel AO (2000) Skin color measurements: comparison between three instruments: the Chromameter(R), the DermaSpectrometer(R) and the Mexameter(R). *Skin Res Technol* 6(4):230-238.
33. Van den Kerckhove E, Staes F, Flour M, Stappaerts K, & Boeckx W (2001) Reproducibility of repeated measurements on healthy skin with Minolta Chromameter CR-300. *Skin Res Technol* 7(1):56-59.
34. Klingenberg CP (2008) MorphoJ (Faculty of Life Sciences, University of Manchester, UK. http://www.flywings.org.uk/MorphoJ_page.htm).
35. Wiley DF, N. Amenta, D. A. Alcantara, D. Ghosh, Y. J. Kil, E. Delson, W. Harcourt-Smith, F. J. Rohlf, K. St. John, and B. Hamann. 2005. (2005) Evolutionary morphing. *Proceedings of the IEEE Visualization 2005* (Proceedings of the IEEE Visualization 2005).
36. Paschou P, Lewis J, Javed A, & Drineas P (2010) Ancestry informative markers for fine-scale individual assignment to worldwide populations. *J. Med. Genet.* 47(12):835-847.
37. Rosenberg NA, Li LM, Ward R, & Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* 73(6):1402-1422.

38. Li JZ, *et al.* (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100-1104.
39. Reich D, *et al.* (2012) Reconstructing Native American population history. *Nature* 488(7411):370-374.
40. Scharf JM, *et al.* (2012) Genome-wide association study of Tourette's syndrome. *Mol. Psychiatry*.
41. Alexander DH, Novembre J, & Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655-1664.
42. R_Development_Core_Team (2010) R: A language and environment for statistical computingVienna).
43. Drake AG & Klingenberg CP (2008) The pace of morphological change: historical transformation of skull shape in St Bernard dogs. *Proceedings. Biological sciences / The Royal Society* 275(1630):71-76.
44. Fujimoto A, *et al.* (2008) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* 17(6):835-843.
45. Fujimoto A, *et al.* (2008) A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia. *Hum. Genet.* 124(2):179-185.
46. Mou C, *et al.* (2008) Enhanced ectodysplasin-A receptor (EDAR) signaling alters multiple fiber characteristics to produce the East Asian hair form. *Hum. Mutat.* 29(12):1405-1411.
47. Relethford JH (2002) Apportionment of global human genetic diversity based on craniometrics and skin color. *Am. J. Phys. Anthropol.* 118(4):393-398.
48. Jablonski NG (2012) Human skin pigmentation as an example of adaptive evolution. *Proc Am Philos Soc* 156(1):45-57.
49. Jablonski NG & Chaplin G (2012) Human skin pigmentation, migration and disease susceptibility. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367(1590):785-792.
50. Sabeti PC, *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164):913-918.
51. Williamson SH, *et al.* (2007) Localizing recent adaptive evolution in the human genome. *PLoS.Genet.* 3(6):e90.
52. (IBGE) IBdGeE (2000) *Brasil: 500 anos de povoamento.* (IBGE, Rio de Janeiro).
53. Curtin PD (1969) *The Atlantic slave trade: a census* (University of Wisconsin Press, Madison).
54. Thomas H (1997) *The Slave Trade* (Simon and Schuster).
55. Safford F & Palacios M (2001) *Colombia: Fragmented Land, Divided Society (Latin American Histories)* (Oxford University Press).
56. Codazzi IGA (2002) *Atlas de Colombia* (IGAC, Bogota, Colombia).
57. Jablonski NG (2006) *Skin: A Natural History* (University of California Press, Berkeley).
58. Sturm RA (2009) Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* 18(R1):R9-17.
59. Kayser M, *et al.* (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet* 82(2):411-423.
60. Liu F, *et al.* (2010) Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet* 6:e1000934.
61. Valverde P, Healy E, Jackson I, Rees JL, & Thody AJ (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat. Genet.* 11(3):328-330.
62. Ganesan AK, *et al.* (2008) Genome-wide siRNA-based functional genomics of pigmentation identifies novel genes and pathways that impact melanogenesis in human cells. *PLoS Genet* 4(12):e1000298.

63. Medland SE, Zhu G, & Martin NG (2009) Estimating the heritability of hair curliness in twins of European ancestry. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 12(5):514-518.
64. Berardesca E, Leveque J-L, & Maibach HI (2007) *Ethnic skin and hair* (Informa Healthcare, New York) pp xii, 269 p.
65. Schmidt-Ullrich R & Paus R (2005) Molecular principles of hair follicle induction and morphogenesis. *Bioessays* 27(3):247-261.
66. Carlson CS*, et al.* (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 15(11):1553-1565.
67. Williamson SH*, et al.* (2005) Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 102(22):7882-7887.
68. Medland SE*, et al.* (2009) Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 85(5):750-755.
69. Nishimura EK, Granter SR, & Fisher DE (2005) Mechanisms of hair graying: incomplete melanocyte stem cell maintenance in the niche. *Science* 307(5710):720-724.
70. Arck PC*, et al.* (2006) Towards a "free radical theory of graying": melanocyte apoptosis in the aging human hair follicle is an indicator of oxidative stress induced tissue damage. *FASEB J.* 20(9):1567-1569.
71. Schouwey K & Beermann F (2008) The Notch pathway: hair graying and pigment cell homeostasis. *Histol. Histopathol.* 23(5):609-619.
72. Vijg J & Campisi J (2008) Puzzles, promises and a cure for ageing. *Nature* 454(7208):1065-1071.
73. Ali G, Chishti MS, Raza SI, John P, & Ahmad W (2007) A mutation in the lipase H (LIPH) gene underlie autosomal recessive hypotrichosis. *Hum. Genet.* 121(3-4):319-325.
74. Azeem Z*, et al.* (2008) Novel mutations in G protein-coupled receptor gene (P2RY5) in families with autosomal recessive hypotrichosis (LAH3). *Hum. Genet.* 123(5):515-519.
75. Tobin DJ & Paus R (2001) Graying: gerontobiology of the hair follicle pigmentary unit. *Exp. Gerontol.* 36(1):29-54.
76. Van Neste D & Tobin DJ (2004) Hair cycle and hair pigmentation: dynamic interactions and changes associated with aging. *Micron* 35(3):193-200.
77. Hillmer AM*, et al.* (2008) Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nat. Genet.* 40(11):1279-1281.
78. Li R*, et al.* (2012) Six novel susceptibility Loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS Genet* 8(5):e1002746.
79. Keogh EV & Walsh RJ (1965) Rate of greying of human hair. *Nature* 207(999):877-878.
80. Cavalli-Sforza LL, Menozzi P, & Piazza A (1994) *The History and Geography of Human Genes* (Princeton U.P., Princeton).
81. Gray J (2000) *Human hair diversity* (Blackwell Science, Abingdon, Oxon) p 70 p.
82. Yang J*, et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42(7):565-569.
83. Lango Allen H*, et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317):832-838.
84. McQuillan R*, et al.* (2012) Evidence of inbreeding depression on human height. *PLoS Genet* 8(7):e1002655.
85. Meisel A & Vega M (2007) The biological standard of living (and its convergence) in Colombia, 1870-2003. A tropical success story. *Economics and human biology* 5(1):100-122.
86. Salvatore RD, Coatsworth JH, & Challú AlE (2010) *Living standards in Latin American history : height, welfare, and development, 1750-2000* (Harvard University David Rockefeller Center for Latin American Studies : Distributed by Harvard University Press, Cambridge, Mass.) pp iii, 313 p.

87. Carson EA (2006) Maximum likelihood estimation of human craniometric heritabilities. *Am. J. Phys. Anthropol.* 131(2):169-180.
88. Hennessy RJ & Stringer CB (2002) Geometric morphometric study of the regional variation of modern human craniofacial form. *Am. J. Phys. Anthropol.* 117(1):37-48.
89. Manica A, Amos W, Balloux F, & Hanihara T (2007) The effect of ancient population bottlenecks on human phenotypic variation. *Nature* 448(7151):346-348.
90. Tapadia MD, Cordero DR, & Helms JA (2005) It's all in your head: new insights into craniofacial development and deformation. *J. Anat.* 207(5):461-477.
91. Tassabehji M, *et al.* (2005) GTF2IRD1 in craniofacial development of humans and mice. *Science* 310(5751):1184-1187.
92. Tobin JL, *et al.* (2008) Inhibition of neural crest migration underlies craniofacial dysmorphology and Hirschsprung's disease in Bardet-Biedl syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 105(18):6714-6719.
93. Liu J, *et al.* (1996) Genetic Analysis of a Morphological Shape Difference in the Male Genitalia of Drosophila simulans and D. mauritiana. *Genetics.* 142:1129-1145.
94. Laurie CC, True JR, Liu J, & Mercer JM (1997) An introgression analysis of quantitative trait loci that contribute to a morphological difference between Drosophila simulans and D. mauritiana. *Genetics* 145(2):339-348.
95. Zeng ZB, *et al.* (2000) Genetic architecture of a morphological shape difference between two Drosophila species. *Genetics* 154(1):299-310.
96. Zimmerman E, Palsson A, & Gibson G (2000) Quantitative trait loci affecting components of wing shape in Drosophila melanogaster. *Genetics* 155(2):671-683.
97. Klingenberg CP, Leamy LJ, Routman EJ, & Cheverud JM (2001) Genetic architecture of mandible shape in mice: Effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics* 157(2):785-802.
98. Klingenberg CP, Leamy LJ, & Cheverud JM (2004) Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. *Genetics* 166(4):1909-1921.
99. Kimura T, *et al.* (2007) Genetic analysis of craniofacial traits in the medaka. *Genetics* 177(4):2379-2388.
100. Kenney-Hunt JP, *et al.* (2008) Pleiotropic patterns of quantitative trait loci for 70 murine skeletal traits. *Genetics* 178(4):2275-2288.
101. Martinez-Abadias N, *et al.* (2006) Phenotypic evolution of human craniofacial morphology after admixture: a geometric morphometrics approach. *Am.J.Phys.Anthropol.* 129(3):387-398.
102. Sardi ML, Ramirez RF, Gonzalez-Jose R, & Pucciarelli HM (2005) South Amerindian craniofacial morphology: diversity and implications for Amerindian evolution. *Am J.Phys.Anthropol.* 128(4):747-756.
103. Hanihara T (2000) Frontal and facial flatness of major human populations. *Am. J. Phys. Anthropol.* 111(1):105-134.
104. Paternoster L, *et al.* (2012) Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am J Hum Genet* 90(3):478-485.
105. Liu F, *et al.* (2012) A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet* 8(9):e1002932.
106. Martinez-Abadias N, *et al.* (2009) Heritability of human cranial dimensions: comparing the evolvability of different cranial regions. *J. Anat.* 214(1):19-35.
107. Martinez-Abadias N, *et al.* (2011) FGF/FGFR signaling coordinates skull development by modulating magnitude of morphological integration: evidence from Apert syndrome mouse models. *PloS one* 6(10):e26425.
108. Enlow DH (1990) *Facial growth* (Saunders, Philadelphia).

109. Ackermann RR, Cheverud JM, Pigliucci M, & Preston K (2004) Morphological integration in primate evolution. *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*, (Oxford University Press, Oxford), pp 302-319.
110. Lieberman DE, McBratney BM, & Krovitz G (2002) The evolution and development of cranial form in Homo sapiens. *Proc.Natl.Acad.Sci.* 99(3):1134-1139.
111. Lieberman DE, Krovitz GE, & McBratney-Owen B (2004) Testing hypotheses about tinkering in the fossil record: the case of the human skull. *J.Exp.Zoolog.B Mol.Dev.Evol.* 302(3):284-301.
112. Bastir M, O'Higgins P, & Rosas A (2007) Facial ontogeny in Neanderthals and modern humans. *Proc.Biol.Sci.* 274(1614):1125-1132.
113. Martinez-Abadias N, Paschetta, C., de Azevedo, S., Esparza, M., Gonzalez-Jose, R. (2009) Developmental and genetic constraints on neurocranial nlobularity: Insights from analyses of deformed skulls and quantitative genetics. *Evolutionary Biology* 36:37-56.
114. Roseman CC (2004) Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proc.Natl.Acad.Sci.U.S.A* 101(35):12824-12829.
115. Roseman CC & Weaver TD (2004) Multivariate apportionment of global human craniometric diversity. *Am. J. Phys. Anthropol.* 125(3):257-263.
116. Gonzalez-Jose RG, *et al.* (2005) Functional-cranial approach to the influence of economic strategy on skull morphology. *Am. J. Phys. Anthropol.* 128(4):757-771.
117. Harvati K & Weaver TD (2006) Human cranial anatomy and the differential preservation of population history and climate signatures. *The anatomical record. Part A, Discoveries in molecular, cellular, and evolutionary biology* 288(12):1225-1233.
118. Florez JC, *et al.* (2009) Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. *Diabetologia* 52(8):1528-1536.
119. Campbell DD, *et al.* (2012) Amerind ancestry, socioeconomic status and the genetics of type 2 diabetes in a Colombian population. *PloS one* 7(4):e33570.
120. Mörner M, Cornell University. Latin American Studies Program., & Columbia University. Institute of Latin American Studies. (1970) *Race and class in Latin America* (Columbia University Press, New York,) pp viii, 309 p.
121. Burkholder MA & Johnson LL (2003) *Colonial Latin America* (Oxford University Press).

Table 1. Sample size, proportion of women, age (median) and estimated genetic ancestry proportion (median) for the sample of each country and in the total sample.

	Brazil	Chile	Colombia	Mexico	Peru	Total
N	1,600	1,561	1,671	1,652	454	6,938
Women	0.67	0.34	0.56	0.60	0.61	0.55
Age	25	24	23	25	20	24
America	0.09	0.48	0.29	0.56	0.64	0.37
Africa	0.09	0.04	0.11	0.05	0.04	0.07
Europe	0.81	0.48	0.61	0.39	0.32	0.56

Table 2. Median of anthropometric features and skin pigmentation in females and males (F/M) of the samples studied

	Brazil	Chile	Colombia	Mexico	Peru	Total
Weight (kg)	60/76	61/76	56/70	60/74	56/71	59/74
Height (cm)	162/175	159/172	160/173	158/172	158/171	160/172
BMI (kg/cm^2)	23/25	25/26	22/23	24/25	23/24	23/25
Hip circum. (cm)	98/100	100/102	94/95	95/97	96/99	97/99
Waist circum. (cm)	75/87	77/90	77/81	81/87	80/88	78/87
ICO	0.46/0.50	0.48/0.52	0.48/0.47	0.51/0.51	0.51/0.52	0.49/0.51
Head circum. (cm)	55/57	56/57	54/56	55/57	55/57	55/57
MI	32/32	36/35	34/33	36/35	37/37	34/35

Note: BMI = Body Mass Index; ICO = Index of Central Obesity; MI = Melanin Index

Table 3. Regression of phenotypes on genetic estimates of European and African ancestry.

Trait	European ancestry Coef.	European ancestry P-value	African ancestry Coef.	African ancestry P-value	R^2	Diff. R^2
Balding	0.16	***1.99E-04***	0.17	0.04	0.19	0.00
Greying	0.10	8.82E-03	0.09	0.19	0.53	0.00
Hair shape	0.49	***5.29E-15***	2.43	***2.00E-16***	0.13	**0.10**
Hair color	-0.58	***2.00E-16***	-0.03	0.67	0.11	**0.05**
Eye color	-1.11	***2.00E-16***	0.25	0.14	0.10	**0.04**
MI	-9.18	***2.00E-16***	12.88	***2.00E-16***	0.25	**0.19**
Weight	0.15	0.88	4.58	0.02	0.33	0.01
Height	6.77	***2.00E-16***	8.03	***5.05E-16***	0.54	0.01
BMI	-1.92	***5.43E-09***	-0.61	0.32	0.18	0.01
Waist circum.	-4.22	***2.05E-06***	-3.20	0.05	0.27	0.01
Hip circum.	-0.20	0.78	2.56	0.06	0.12	0.01
ICO	-0.05	***2.00E-16***	-0.04	***9.63E-06***	0.23	**0.02**
Head circum.	-0.26	0.28	0.45	0.33	0.26	0.01
CS	-12.98	***7.74E-05***	2.94	0.62	0.48	0.00
PC1(19%)	0.02	***1.42E-11***	0.04	***1.24E-12***	0.15	**0.02**
PC2(12%)	-0.01	1.00E-03	-0.02	***7.80E-06***	0.02	0.00
PC3(10%)	-0.01	***8.85E-13***	0.00	0.84	0.21	**0.03**
PC4(7%)	0.01	***2.34E-08***	-0.02	***6.14E-08***	0.18	**0.05**
PC5(7%)	0.01	***3.16E-05***	0.00	0.60	0.05	0.00
European shape score	0.02	***2.00E-16***	0.01	7.06E-04	0.34	0.26

Note: Being proportions, European, African and American ancestries sum up to 1 and since in this sample African ancestry is very low (median of 7%), we use Native American ancestry as a baseline. All regressions account for age, sex, sampling site and SEP (EA and WI). Regressions for facial traits also account for BMI and height. P-Values < 10^{-3} are shown in bold italic. % Ancestry refers to variation explained by incorporating European and African ancestry in the model (with values greater than 1% highlighted in bold). BMI = Body Mass Index; ICO = Index of Central Obesity; MI = Melanin Index; PC = Principal Component (% in parenthesis refers to phenotypic variance explained by that PC); Diff. R^2 refers to the difference in R^2 for the full regression model and a model without European and African ancestry as explanatory variables.

Table 4. Regression of the difference between individual self-perception and genetic ancestry bands.

	Δ AMERICA		Δ EUROPE		Δ AFRICA	
	Coef.	P-value	Coef.	P-value	Coef.	P-value
Age	0.00	0.93	0.00	0.87	0.00	0.16
Sex(M)	-0.05	0.13	0.26	***4.25E-15***	-0.12	***1.87E-07***
Chile	0.18	0.01	-0.61	***2.00E-16***	-0.27	***4.16E-09***
Colombia	0.57	***2.00E-16***	-0.88	***2.00E-16***	-0.09	0.03
Mexico	0.47	***1.29E-11***	-0.51	***3.55E-14***	-0.26	***7.52E-08***
Peru	-0.03	0.78	-0.80	***2.27E-11***	-0.35	***2.43E-05***
European anc.	0.62	***2.00E-16***	-0.75	***2.00E-16***	-0.02	0.15
Africa anc.	0.46	***2.00E-16***	0.06	0.22	-0.60	***2.00E-16***
EA	0.04	0.04	0.05	0.01	-0.02	0.18
MI	0.02	***1.68E-07***	-0.05	***2.00E-16***	0.03	***2.00E-16***
Hair color	0.10	4.84E-03	-0.17	***2.86E-06***	0.05	0.05
Eye color	0.08	***2.52E-06***	-0.11	***4.55E-10***	0.01	0.27
Hair shape	-0.06	0.01	0.01	0.66	0.15	***2.00E-16***

NOTE: Δ refers to self-perception (bands 1 to 5) minus continental ancestry (grouped into 1 to 5 bands at 20 intervals) estimated from the genetic data. MI= Melanin index, EA = Educational Attainment. Sex and Sampling site (here with country names) were incorporated in the analyses as factors while the other variables were treated as quantitative.

Figure 1. Frequency distribution of individual ancestry values in each country.

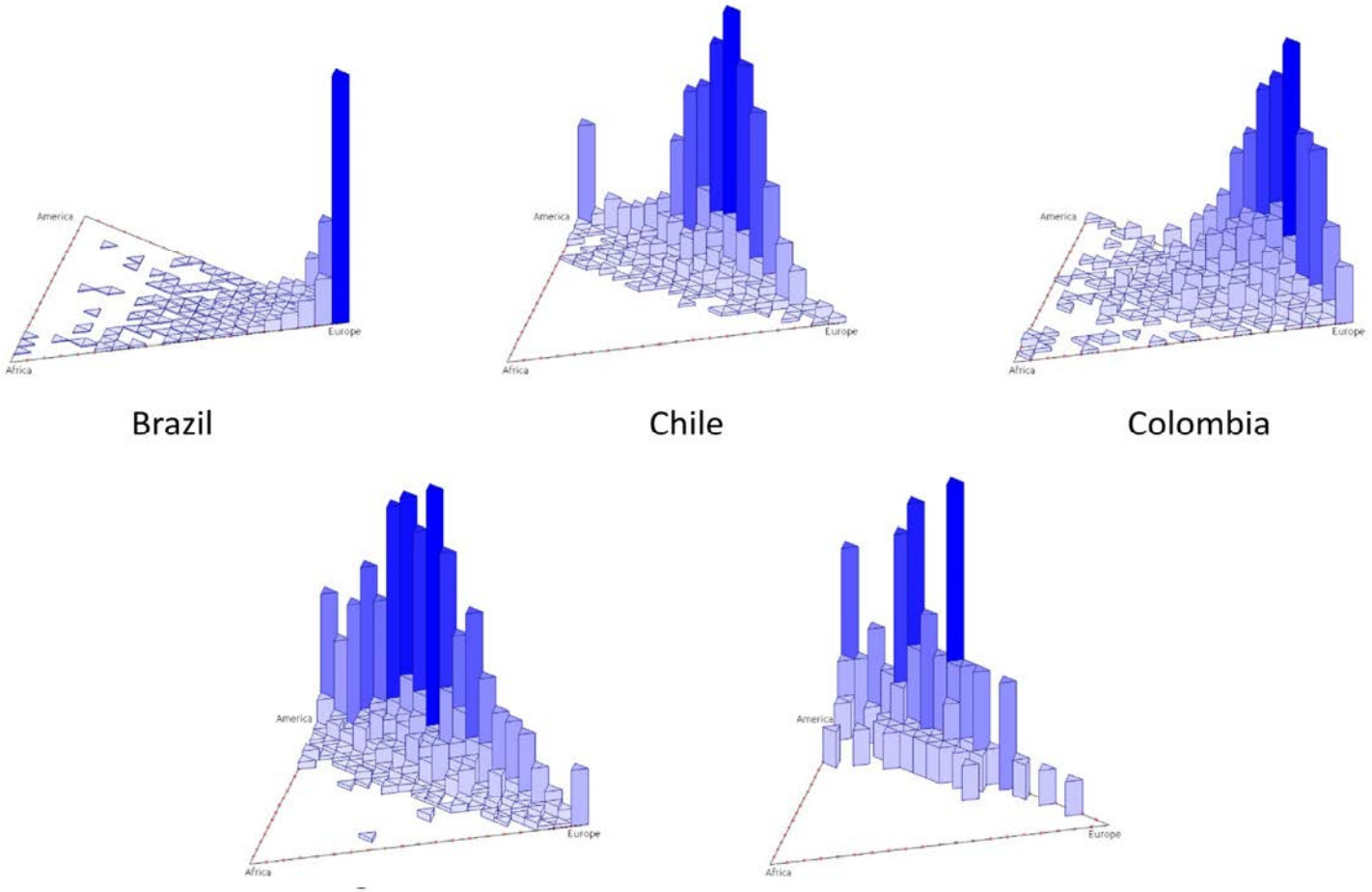


Figure 2. Geographic distribution of individual ancestry values in each country.

Fig 2A

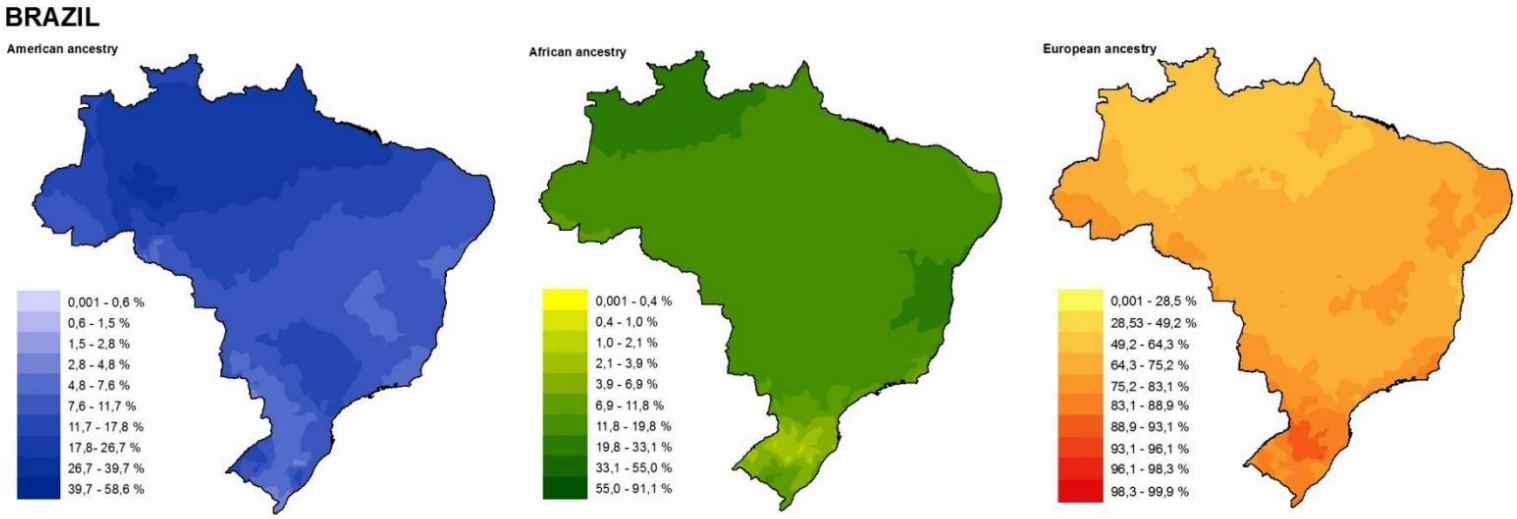


Fig 2B

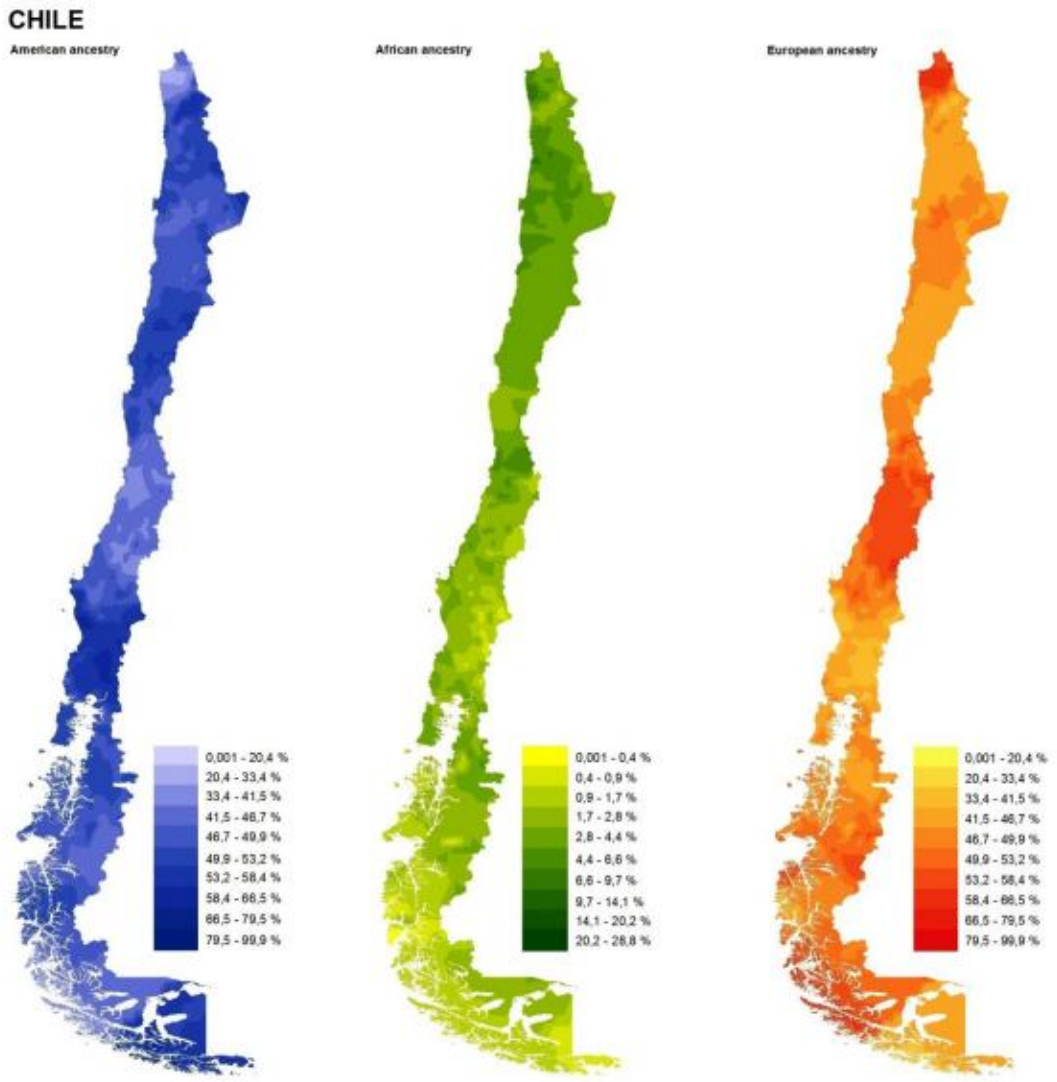

CHILE
American ancestry
0,001 - 20,4 %
20,4 - 33,4 %
33,4 - 41,5 %
41,5 - 46,7 %
46,7 - 49,9 %
49,9 - 53,2 %
53,2 - 58,4 %
58,4 - 66,5 %
66,5 - 79,5 %
79,5 - 99,9 %
African ancestry
0,001 - 0,4 %
0,4 - 0,9 %
0,9 - 1,7 %
1,7 - 2,8 %
2,8 - 4,4 %
4,4 - 6,6 %
6,6 - 9,7 %
9,7 - 14,1 %
14,1 - 20,2 %
20,2 - 28,8 %
European ancestry
0,001 - 20,4 %
20,4 - 33,4 %
33,4 - 41,5 %
41,5 - 46,7 %
46,7 - 49,9 %
49,9 - 53,2 %
53,2 - 58,4 %
58,4 - 66,5 %
66,5 - 79,5 %
79,5 - 99,9 %

Fig 2C

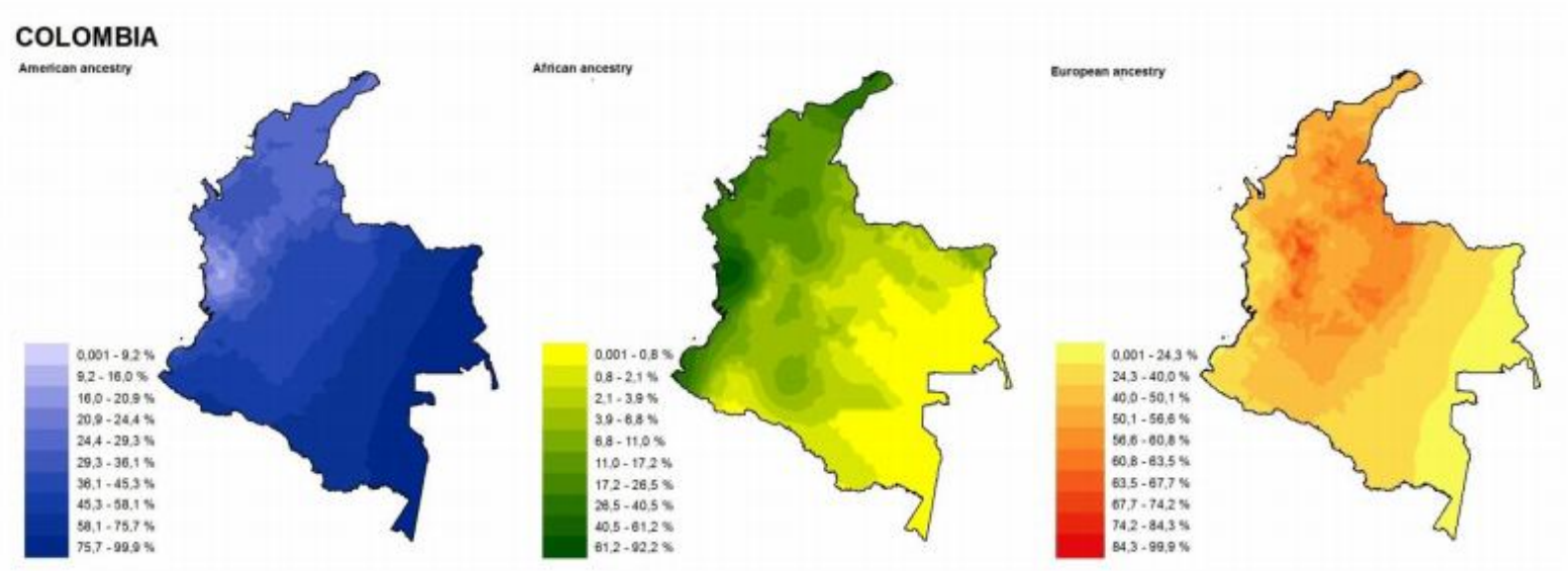
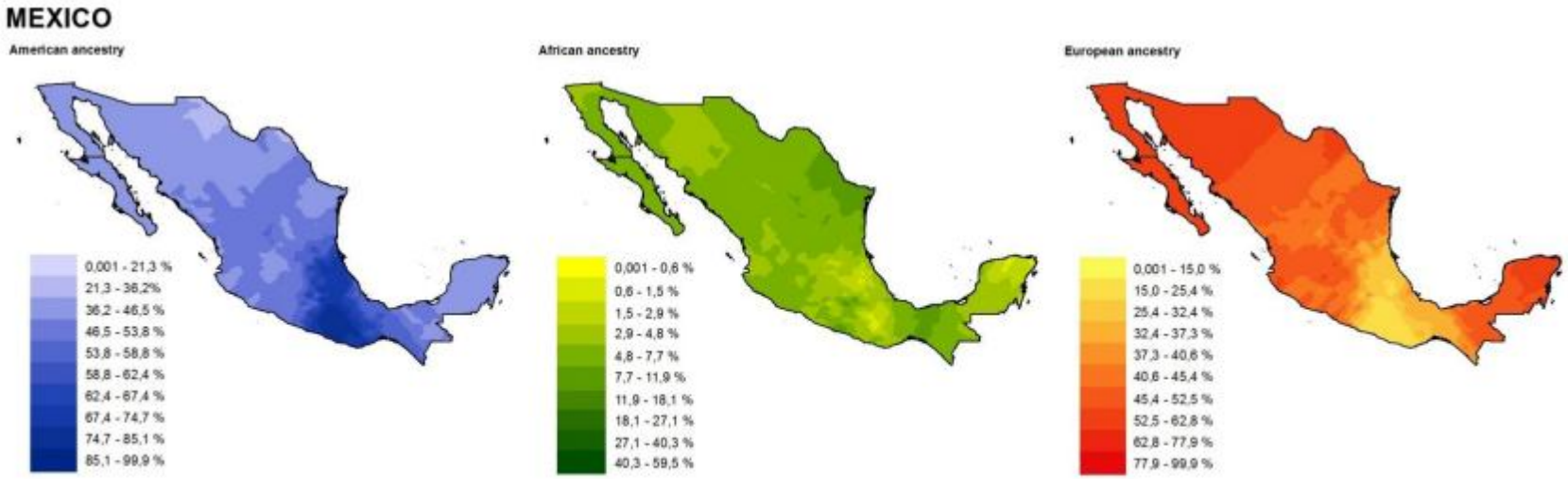

COLOMBIA
American ancestry
0,001 - 9,2 %
9,2 - 16,0 %
16,0 - 20,9 %
20,9 - 24,4 %
24,4 - 29,3 %
29,3 - 36,1 %
36,1 - 45,3 %
45,3 - 58,1 %
58,1 - 75,7 %
75,7 - 99,9 %
African ancestry
0,001 - 0,8 %
0,8 - 2,1 %
2,1 - 3,9 %
3,9 - 6,8 %
6,8 - 11,0 %
11,0 - 17,2 %
17,2 - 26,5 %
26,5 - 40,5 %
40,5 - 61,2 %
61,2 - 92,2 %
European ancestry
0,001 - 24,3 %
24,3 - 40,0 %
40,0 - 50,1 %
50,1 - 56,6 %
56,6 - 60,8 %
60,8 - 63,5 %
63,5 - 67,7 %
67,7 - 74,2 %
74,2 - 84,3 %
84,3 - 99,9 %

Fig 2D


MEXICO
American ancestry
0,001 - 21,3 %
21,3 - 36,2%
36,2 - 46,5 %
46,5 - 53,8 %
53,8 - 58,8 %
58,8 - 62,4 %
62,4 - 67,4 %
67,4 - 74,7 %
74,7 - 85,1 %
85,1 - 99,9 %
African ancestry
0,001 - 0,6 %
0,6 - 1,5 %
1,5 - 2,9 %
2,9 - 4,8 %
4,8 - 7,7 %
7,7 - 11,9 %
11,9 - 18,1 %
18,1 - 27,1 %
27,1 - 40,3 %
40,3 - 59,5 %
European ancestry
0,001 - 15,0 %
15,0 - 25,4 %
25,4 - 32,4 %
32,4 - 37,3 %
37,3 - 40,6 %
40,6 - 45,4 %
45,4 - 52,5 %
52,5 - 62,8 %
62,8 - 77,9 %
77,9 - 99,9 %

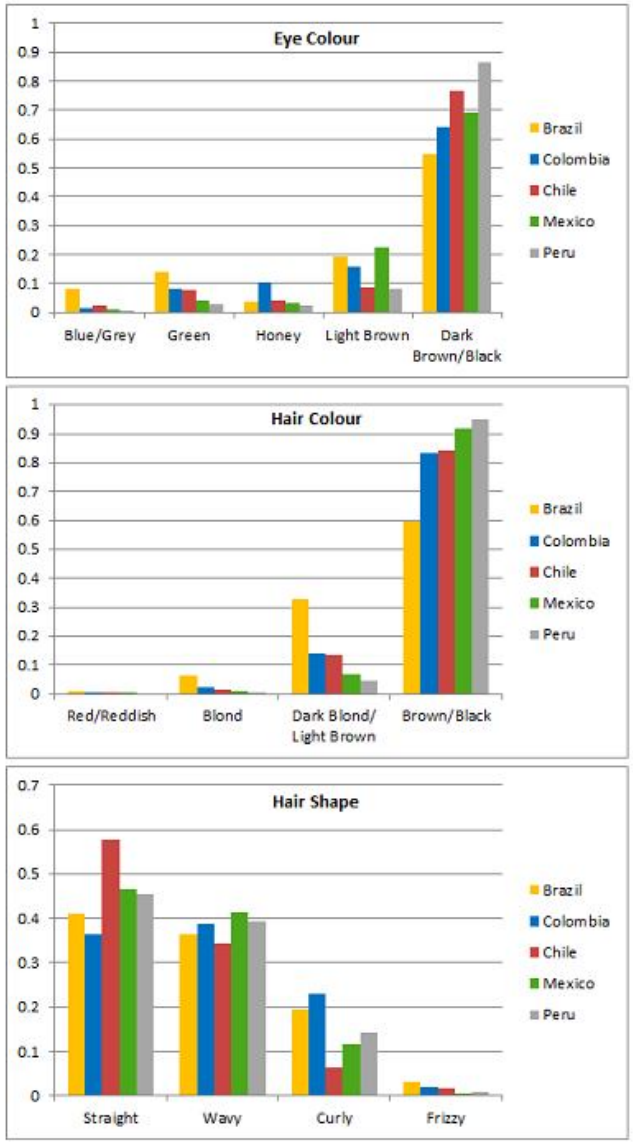


Figure 3. Distribution of eye color, hair color and hair type in each country.

Figure 4. Box plots relating hair shape, hair color and eye color categories to European ancestry in the total sample.

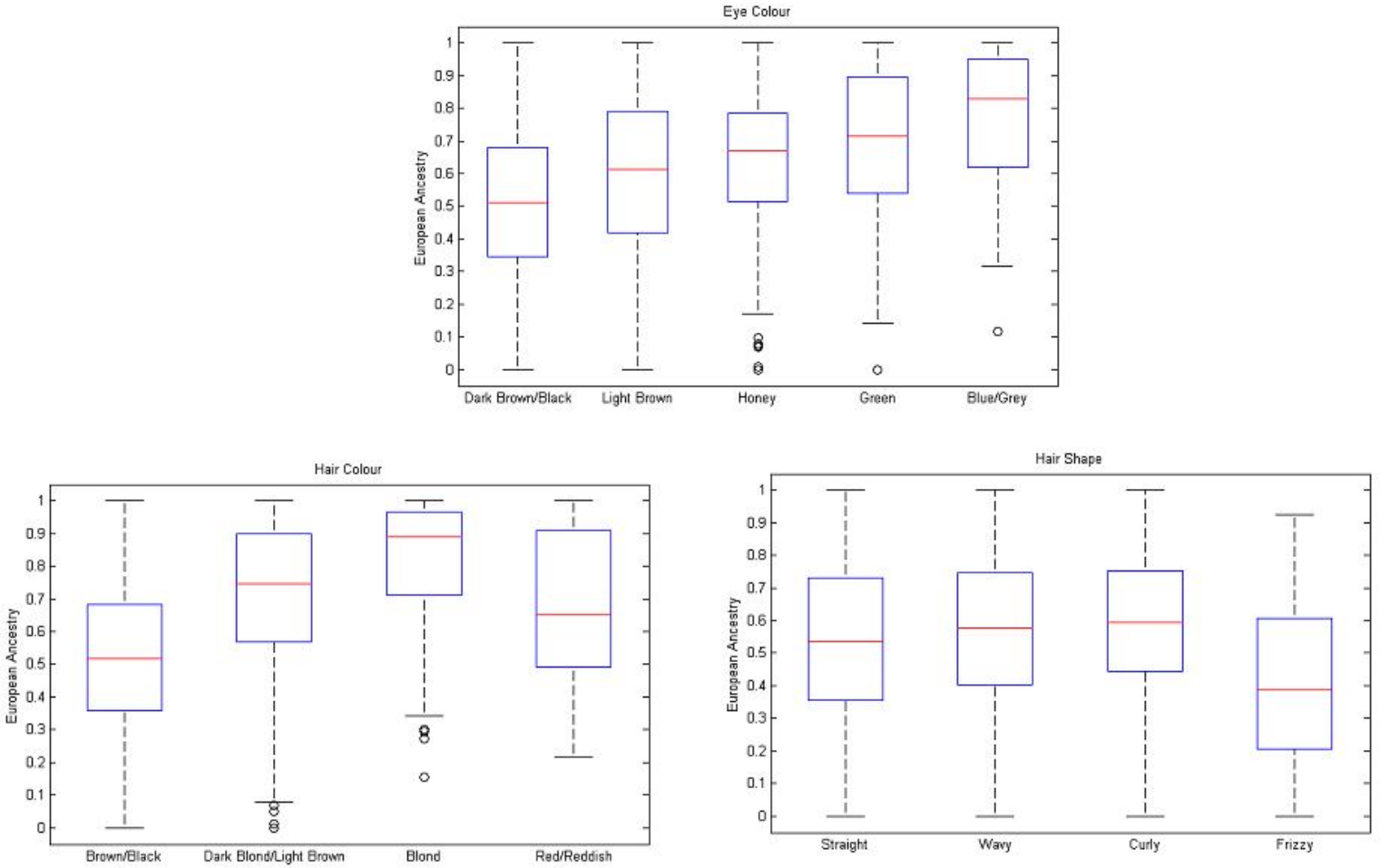


Figure 5. Scatter plots relating (A) skin pigmentation (melanin index MI) (B) height and (C) Shape score, and European ancestry in the total study sample.

Fig 5A

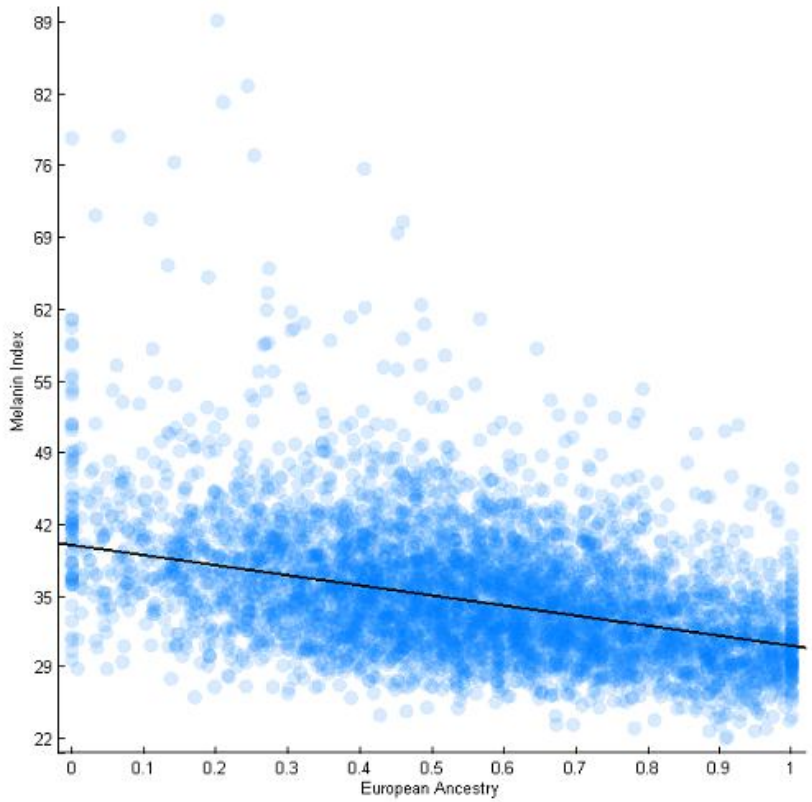


Fig5B

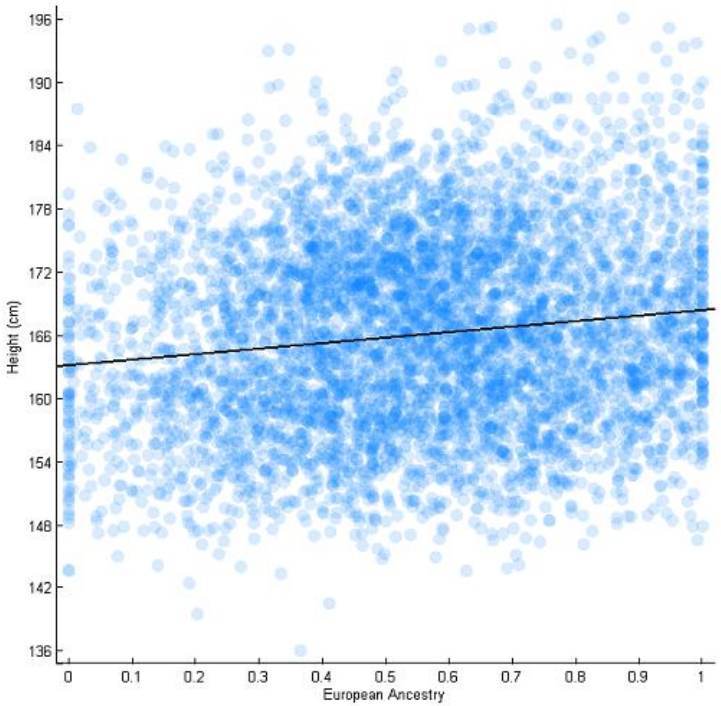

Height (cm)
196
190
184
178
172
166
160
154
148
142
136
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
European Ancestry

Fig 5C

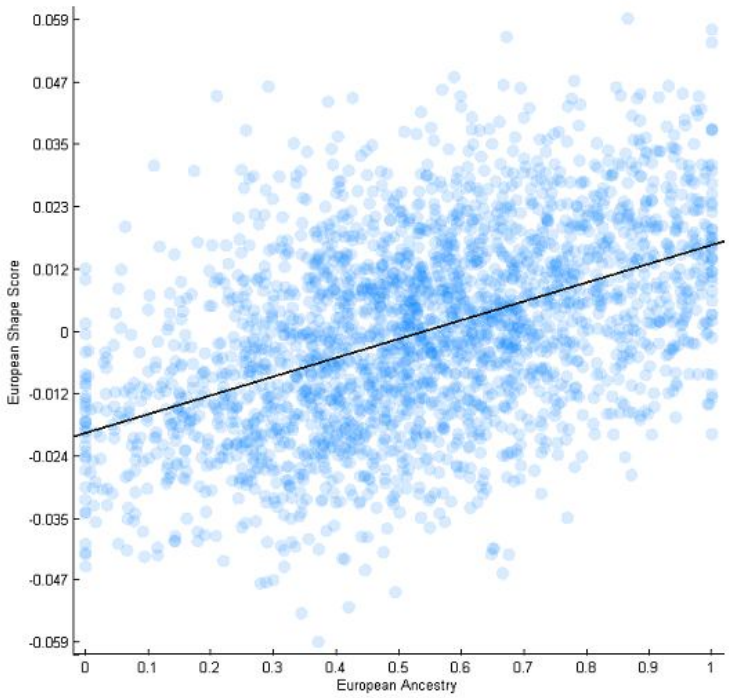
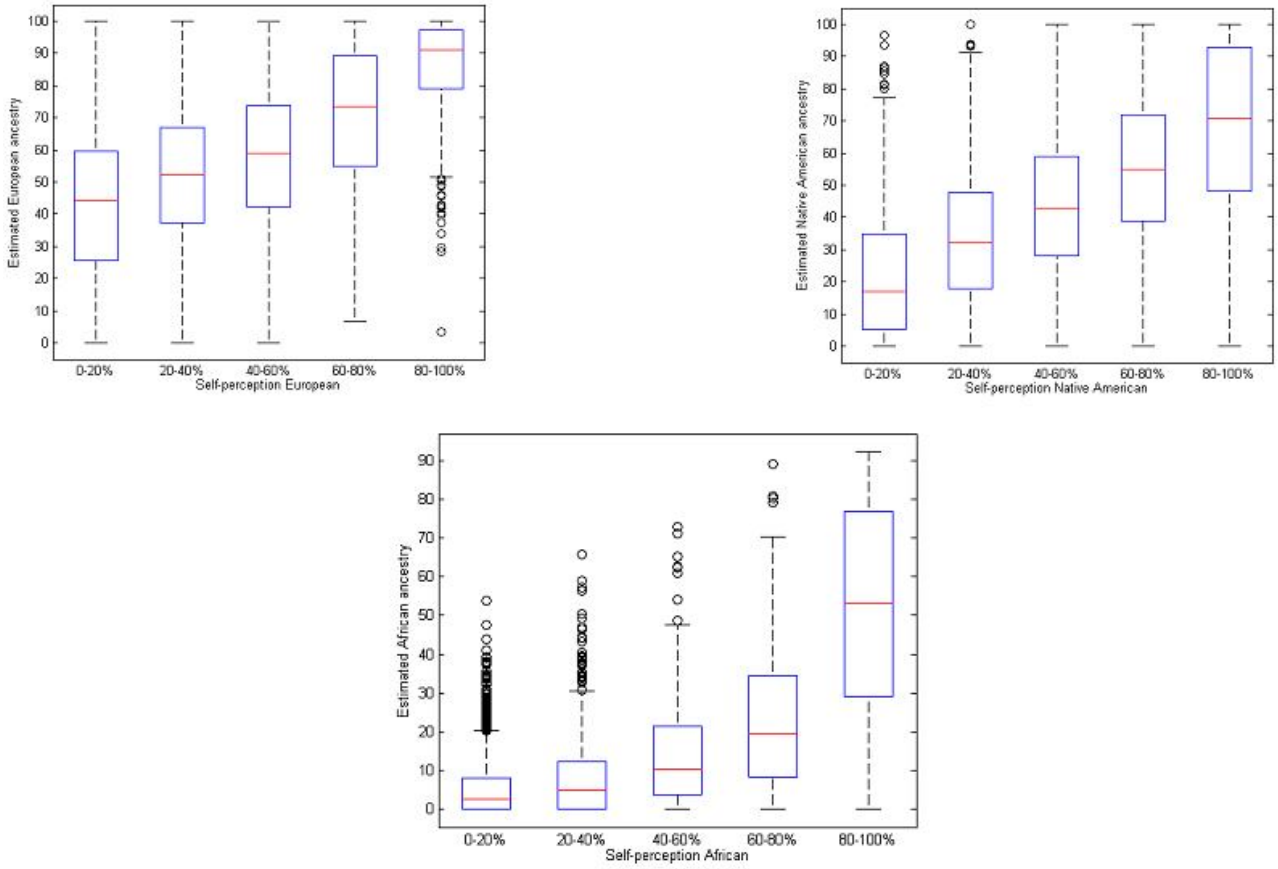

European Shape Score
0.059
0.047
0.035
0.023
0.012
0
-0.012
-0.024
-0.035
-0.047
-0.059
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
European Ancestry

Figure 6. Box plots relating individual genetic ancestry estimates to the bands of self-perceived African, Native American and European ancestry.



Supplementary Information

Table of Contents

Supplementary Notes

Supplementary Note S1: Indicators of socioeconomic position (SEP).

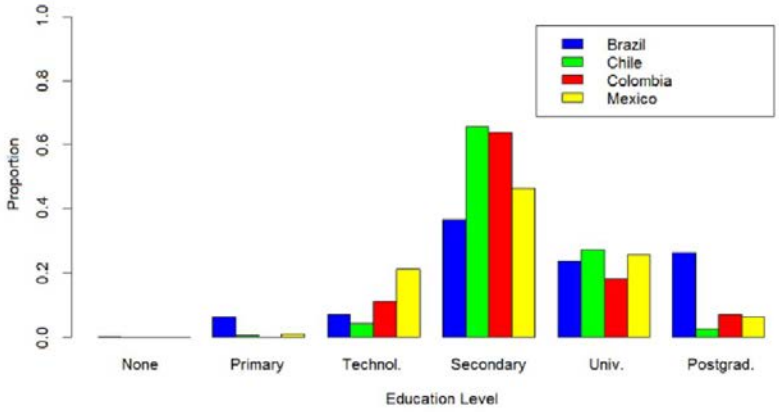
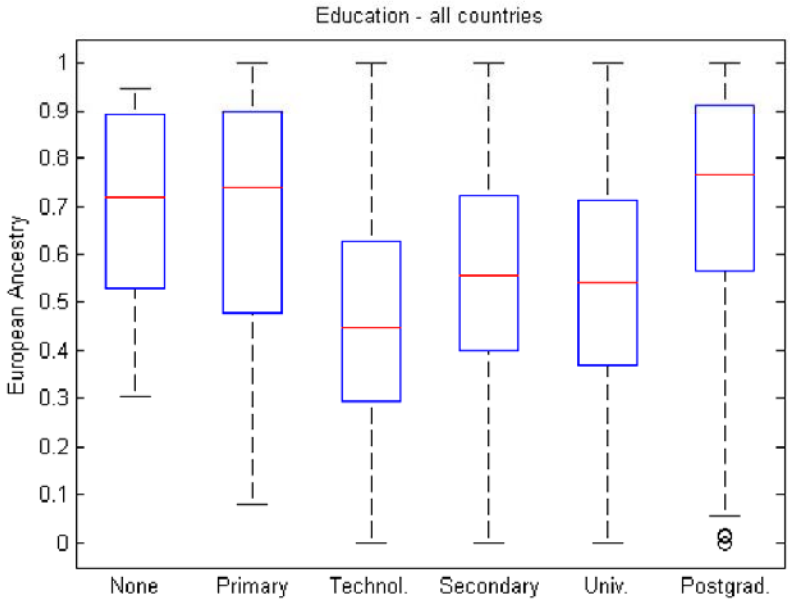
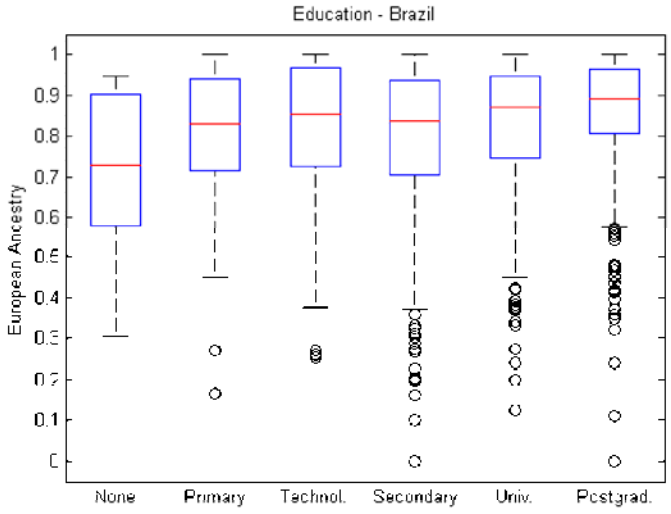
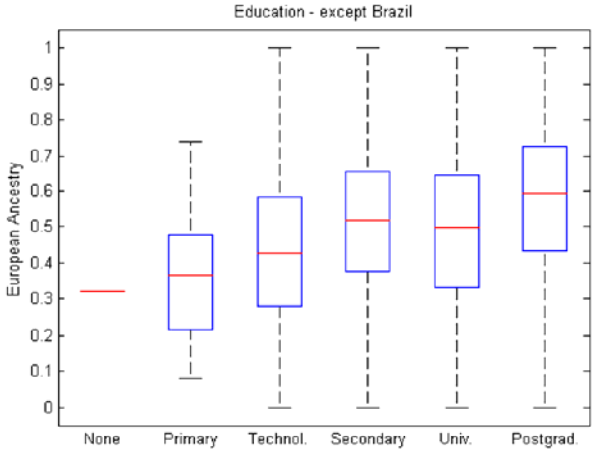
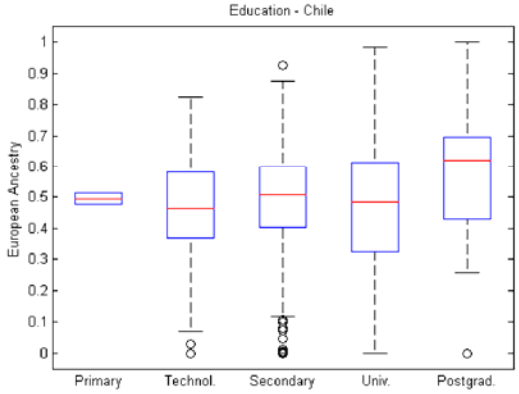


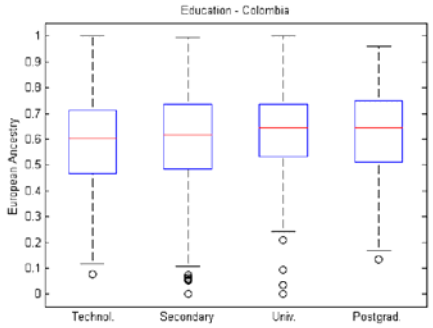
Figure 1 Distribution of maximal educational attainment by country sample

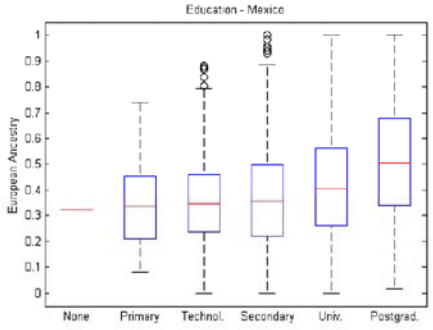



Education - Brazil
European Ancestry
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
None
Primary
Technol.
Secondary
Univ.
Postgrad.


Education - except Brazil
European Ancestry
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
None
Primary
Technol.
Secondary
Univ.
Postgrad.


Education - Chile
European Ancestry
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Primary
Technol.
Secondary
Univ.
Postgrad.



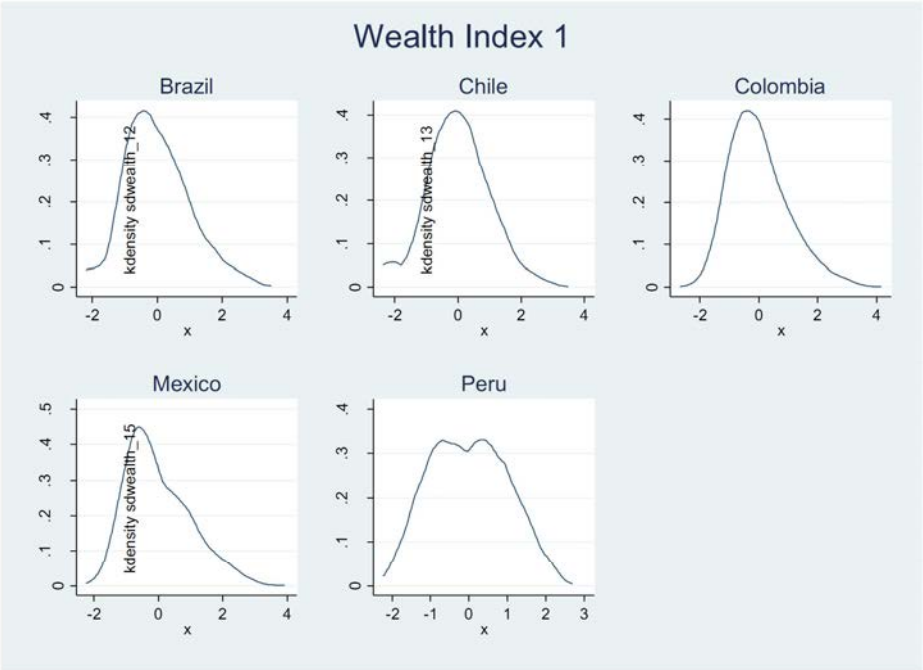


Construction of Wealth Index:

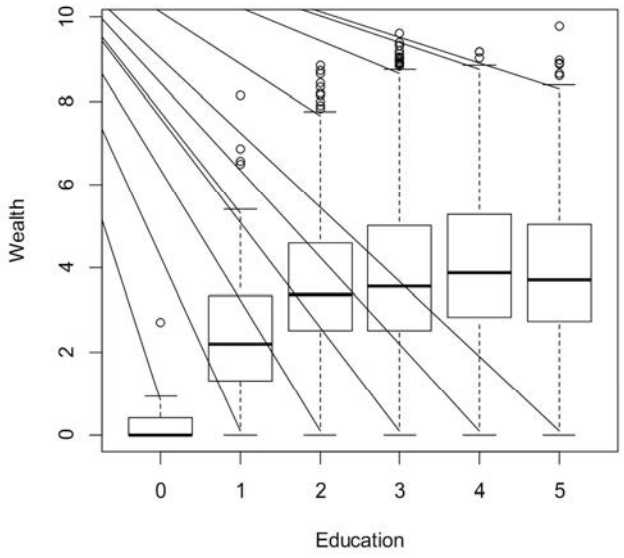
Number of Observations by Country -- the first column third column 'N (analysis)' reports the effective number of observations used for the construction of country-specific wealth indexes. We dropped observations (individuals) for which more than 7 of the assets used in the construction of the indexes were missing.

	N (original)	>7 assets missing	N (analysis)
Brazil	1,600	121	1479
Chile	1,561	34	1527
Colombia	1,678	42	1636
Mexico	1,682	20	1662
Peru	454	43	411

The index was standardised. The distribution of this standardised index by country is shown in the figure below.



Boxplot comparing the wealth index for different education categories.



Supplementary Note S2: Regression of phenotypes on ancestry and covariates.

Weight(Kg)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     51.66008    1.22776  42.077  < 2e-16 ***
AGE              0.25780    0.01727  14.927  < 2e-16 ***
SEXM            14.18079    0.34885  40.650  < 2e-16 ***
COUNTRYCHILE     1.12957    0.66650   1.695   0.0902 .
COUNTRYCOLOMBIA -4.10125    0.53528  -7.662 2.19e-14 ***
COUNTRYMEXICO   -0.54525    0.69313  -0.787   0.4315
COUNTRYPERU     -3.25960    1.44588  -2.254   0.0242 *
EUROPE           0.15145    1.01473   0.149   0.8814
AFRICA           4.57987    1.88994   2.423   0.0154 *
nbEDUCATION      0.36305    0.19314   1.880   0.0602 .
wealth_1         0.47821    0.10407   4.595 4.43e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.68 on 5011 degrees of freedom
  (1916 observations deleted due to missingness)
Multiple R-squared: 0.331, Adjusted R-squared: 0.3296
F-statistic: 247.9 on 10 and 5011 DF,  p-value: < 2.2e-16
```

Height (cm)

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     156.134043   0.640769 243.667  < 2e-16 ***
AGE              -0.098566   0.009014 -10.935  < 2e-16 ***
SEXM             13.159858   0.182064  72.281  < 2e-16 ***
COUNTRYCHILE     -1.249052   0.347845  -3.591 0.000333 ***
COUNTRYCOLOMBIA  -1.825470   0.279363  -6.534 7.02e-11 ***
COUNTRYMEXICO    -0.751177   0.361746  -2.077 0.037896 *
COUNTRYPERU      -0.764645   0.754601  -1.013 0.310961
EUROPE            6.770971   0.529589  12.785  < 2e-16 ***
AFRICA            8.026433   0.986356   8.137 5.05e-16 ***
nbEDUCATION       0.733776   0.100801   7.279 3.87e-13 ***
wealth_1          0.145851   0.054312   2.685 0.007267 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.094 on 5011 degrees of freedom
  (1916 observations deleted due to missingness)
Multiple R-squared: 0.5437, Adjusted R-squared: 0.5428
F-statistic: 597.1 on 10 and 5011 DF,  p-value: < 2.2e-16
```

Hip circumference (cm)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     93.03469    0.88997 104.537  < 2e-16 ***
AGE              0.15128    0.01254  12.066  < 2e-16 ***
SEXM             1.36389    0.25288   5.393 7.23e-08 ***
COUNTRYCHILE     1.89605    0.48309   3.925 8.80e-05 ***
COUNTRYCOLOMBIA -4.19687    0.38806 -10.815  < 2e-16 ***
COUNTRYMEXICO   -2.38709    0.50242  -4.751 2.08e-06 ***
COUNTRYPERU     -2.65457    1.04794  -2.533   0.0113 *
EUROPE          -0.20410    0.73551  -0.277   0.7814
AFRICA           2.55850    1.36998   1.868   0.0619 .
nbEDUCATION      0.16484    0.14005   1.177   0.2393
wealth_1         0.33948    0.07544   4.500 6.94e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.463 on 5009 degrees of freedom
  (1918 observations deleted due to missingness)
Multiple R-squared: 0.1217,Adjusted R-squared:  0.12
F-statistic: 69.41 on 10 and 5009 DF,  p-value: < 2.2e-16
```

Waist circumference (cm)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     71.65787    1.07503  66.656  < 2e-16 ***
AGE              0.38638    0.01515  25.510  < 2e-16 ***
SEXM             7.40481    0.30549  24.239  < 2e-16 ***
COUNTRYCHILE     2.72162    0.58358   4.664 3.19e-06 ***
COUNTRYCOLOMBIA -1.32306    0.46870  -2.823 0.004779 **
COUNTRYMEXICO    2.17978    0.60693   3.591 0.000332 ***
COUNTRYPERU      5.14077    1.26599   4.061 4.97e-05 ***
EUROPE          -4.22428    0.88848  -4.754 2.05e-06 ***
AFRICA          -3.19669    1.65481  -1.932 0.053447 .
nbEDUCATION     -0.44571    0.16919  -2.634 0.008455 **
wealth_1         0.20418    0.09112   2.241 0.025079 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.22 on 5010 degrees of freedom
  (1917 observations deleted due to missingness)
Multiple R-squared: 0.2729,Adjusted R-squared: 0.2715
F-statistic: 188.1 on 10 and 5010 DF,  p-value: < 2.2e-16
```

Head circumference (cm)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     54.510090   0.300586 181.346   <2e-16 ***
AGE             -0.005378   0.004198  -1.281   0.2003
SEXM             2.230419   0.086071  25.914   <2e-16 ***
COUNTRYCHILE     0.317250   0.162205   1.956   0.0505 .
COUNTRYCOLOMBIA -1.591849   0.132743 -11.992   <2e-16 ***
COUNTRYMEXICO   -0.174851   0.168807  -1.036   0.3003
COUNTRYPERU     -0.092598   0.350811  -0.264   0.7918
EUROPE          -0.143440   0.248099  -0.578   0.5632
AFRICA           0.863371   0.469761   1.838   0.0661 .
nbEDUCATION      0.093656   0.047460   1.973   0.0485 *
wealth_1         0.065052   0.025466   2.554   0.0107 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.828 on 4849 degrees of freedom
  (2078 observations deleted due to missingness)
Multiple R-squared: 0.1957,Adjusted R-squared: 0.194
F-statistic:   118 on 10 and 4849 DF,  p-value: < 2.2e-16
```

Melanin Index (MI)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     41.642582   0.545687  76.312  < 2e-16 ***
AGE             -0.011698   0.007683  -1.522 0.127958
SEXM            -0.405672   0.155140  -2.615 0.008953 **
COUNTRYCHILE     0.644998   0.296257   2.177 0.029515 *
COUNTRYCOLOMBIA -0.236020   0.238015  -0.992 0.321432
COUNTRYMEXICO   -0.303512   0.308098  -0.985 0.324616
COUNTRYPERU     -0.684680   0.642308  -1.066 0.286489
EUROPE          -9.178241   0.451095 -20.347  < 2e-16 ***
AFRICA          12.875186   0.840119  15.325  < 2e-16 ***
nbEDUCATION     -0.306057   0.085998  -3.559 0.000376 ***
wealth_1        -0.178486   0.046298  -3.855 0.000117 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.186 on 5000 degrees of freedom
  (1927 observations deleted due to missingness)
Multiple R-squared: 0.2474,Adjusted R-squared: 0.2459
F-statistic: 164.4 on 10 and 5000 DF,  p-value: < 2.2e-16
```

Hairshape

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.245431   0.074968  16.613  < 2e-16 ***
AGE             -0.000923   0.001054  -0.876 0.381240
SEXM             0.015322   0.021298   0.719 0.471915
COUNTRYCHILE    -0.078048   0.040681  -1.919 0.055103 .
```

```
COUNTRYCOLOMBIA  0.152680   0.032675   4.673 3.05e-06 ***
COUNTRYMEXICO    0.164188   0.042305   3.881 0.000105 ***
COUNTRYPERU      0.312461   0.088223   3.542 0.000401 ***
EUROPE           0.485959   0.061951   7.844 5.29e-15 ***
AFRICA           2.427106   0.115284  21.053  < 2e-16 ***
nbEDUCATION     -0.008283   0.011792  -0.702 0.482436
wealth_1        -0.003375   0.006350  -0.531 0.595121
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7123 on 5003 degrees of freedom
  (1924 observations deleted due to missingness)
Multiple R-squared: 0.1303,Adjusted R-squared: 0.1285
F-statistic: 74.92 on 10 and 5003 DF,  p-value: < 2.2e-16
```

Eye color

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.905e+00  1.099e-01  44.612  < 2e-16 ***
AGE             -6.424e-03  1.547e-03  -4.153 3.33e-05 ***
SEXM            -5.562e-02  3.123e-02  -1.781 0.074988 .
COUNTRYCHILE     1.850e-01  5.966e-02   3.102 0.001936 **
COUNTRYCOLOMBIA  1.764e-01  4.795e-02   3.680 0.000236 ***
COUNTRYMEXICO    2.224e-01  6.205e-02   3.584 0.000341 ***
COUNTRYPERU      4.711e-01  1.294e-01   3.640 0.000275 ***
EUROPE          -1.113e+00  9.084e-02 -12.250  < 2e-16 ***
AFRICA           2.490e-01  1.691e-01   1.473 0.140940
nbEDUCATION      2.802e-02  1.729e-02   1.621 0.105179
wealth_1         3.525e-05  9.313e-03   0.004 0.996980
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.045 on 5006 degrees of freedom
  (1921 observations deleted due to missingness)
Multiple R-squared: 0.09632,     Adjusted R-squared: 0.09452
F-statistic: 53.36 on 10 and 5006 DF,  p-value: < 2.2e-16
```

Balding

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.379177   0.052243   7.258 4.53e-13 ***
AGE             0.014905   0.000735  20.279  < 2e-16 ***
SEXM            0.352147   0.014850  23.714  < 2e-16 ***
COUNTRYCHILE    0.080302   0.028350   2.832 0.004637 **
COUNTRYCOLOMBIA 0.048333   0.022773   2.122 0.033855 *
COUNTRYMEXICO   0.108210   0.029487   3.670 0.000245 ***
COUNTRYPERU     0.192669   0.061493   3.133 0.001739 **
EUROPE          0.160809   0.043195   3.723 0.000199 ***
AFRICA          0.166735   0.080368   2.075 0.038071 *
nbEDUCATION     0.005678   0.008224   0.690 0.489992
```

```
wealth_1        0.010950   0.004428   2.473 0.013445 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4965 on 5002 degrees of freedom
  (1925 observations deleted due to missingness)
Multiple R-squared: 0.1899,Adjusted R-squared: 0.1883
F-statistic: 117.3 on 10 and 5002 DF,  p-value: < 2.2e-16
```

Hair color

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.8870733  0.0523357  74.272  < 2e-16 ***
AGE              0.0006871  0.0007356   0.934  0.35036
SEXM             0.0935995  0.0148650   6.297 3.30e-10 ***
COUNTRYCHILE     0.1164509  0.0283985   4.101 4.19e-05 ***
COUNTRYCOLOMBIA  0.2064264  0.0228087   9.050  < 2e-16 ***
COUNTRYMEXICO    0.0667396  0.0295457   2.259  0.02393 *
COUNTRYPERU      0.1728062  0.0616014   2.805  0.00505 **
EUROPE          -0.5790847  0.0432507 -13.389  < 2e-16 ***
AFRICA          -0.0346993  0.0804749  -0.431  0.66635
nbEDUCATION      0.0147293  0.0082337   1.789  0.07369 .
wealth_1        -0.0103336  0.0044342  -2.330  0.01982 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4974 on 5008 degrees of freedom
  (1919 observations deleted due to missingness)
Multiple R-squared: 0.112,  Adjusted R-squared: 0.1102
F-statistic: 63.18 on 10 and 5008 DF,  p-value: < 2.2e-16
```

Graying

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.0112855  0.0460328   0.245  0.80634
AGE              0.0459450  0.0006472  70.994  < 2e-16 ***
SEXM             0.0100926  0.0130769   0.772  0.44028
COUNTRYCHILE     0.0570788  0.0249770   2.285  0.02234 *
COUNTRYCOLOMBIA  0.0413819  0.0200655   2.062  0.03923 *
COUNTRYMEXICO    0.1502173  0.0259802   5.782 7.83e-09 ***
COUNTRYPERU      0.1064898  0.0541825   1.965  0.04942 *
EUROPE           0.0996517  0.0380325   2.620  0.00882 **
AFRICA           0.0936813  0.0707963   1.323  0.18581
nbEDUCATION     -0.0408036  0.0072427  -5.634 1.86e-08 ***
wealth_1         0.0021399  0.0038981   0.549  0.58306
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4376 on 5008 degrees of freedom
```

```
  (1919 observations deleted due to missingness)
Multiple R-squared: 0.5295,Adjusted R-squared: 0.5286
F-statistic: 563.7 on 10 and 5008 DF,  p-value: < 2.2e-16
```

BMI

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     21.6014894  0.3978414  54.297  < 2e-16 ***
AGE              0.1264948  0.0055966  22.602  < 2e-16 ***
SEXM             1.2136129  0.1130404  10.736  < 2e-16 ***
COUNTRYCHILE     0.7545594  0.2159701   3.494 0.000480 ***
COUNTRYCOLOMBIA -0.9716775  0.1734513  -5.602 2.23e-08 ***
COUNTRYMEXICO    0.0003411  0.2246010   0.002 0.998788
COUNTRYPERU     -1.0139647  0.4685174  -2.164 0.030496 *
EUROPE          -1.9214136  0.3288117  -5.844 5.43e-09 ***
AFRICA          -0.6120187  0.6124096  -0.999 0.317668
nbEDUCATION     -0.1152192  0.0625857  -1.841 0.065683 .
wealth_1         0.1217697  0.0337211   3.611 0.000308 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.784 on 5011 degrees of freedom
  (1916 observations deleted due to missingness)
Multiple R-squared: 0.1774,Adjusted R-squared: 0.1758
F-statistic: 108.1 on 10 and 5011 DF,  p-value: < 2.2e-16
```

ICO

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.628e-01  6.383e-03  72.500  < 2e-16 ***
AGE              2.675e-03  8.993e-05  29.749  < 2e-16 ***
SEXM             4.554e-03  1.814e-03   2.511   0.0121 *
COUNTRYCHILE     1.932e-02  3.465e-03   5.576 2.60e-08 ***
COUNTRYCOLOMBIA -2.462e-03  2.783e-03  -0.885   0.3764
COUNTRYMEXICO    1.547e-02  3.604e-03   4.293 1.80e-05 ***
COUNTRYPERU      3.370e-02  7.517e-03   4.484 7.50e-06 ***
EUROPE          -4.608e-02  5.275e-03  -8.736  < 2e-16 ***
AFRICA          -4.352e-02  9.825e-03  -4.430 9.63e-06 ***
nbEDUCATION     -5.315e-03  1.005e-03  -5.291 1.27e-07 ***
wealth_1         6.572e-04  5.410e-04   1.215   0.2245
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06071 on 5010 degrees of freedom
  (1917 observations deleted due to missingness)
Multiple R-squared: 0.2317,Adjusted R-squared: 0.2301
F-statistic: 151.1 on 10 and 5010 DF,  p-value: < 2.2e-16
```

Centroid Size

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     286.24885   14.32954  19.976  < 2e-16 ***
SEXM             15.65232    1.61307   9.703  < 2e-16 ***
AGE              -0.01339    0.06214  -0.216    0.829
COUNTRYCHILE     32.93157    2.16989  15.177  < 2e-16 ***
COUNTRYCOLOMBIA  -0.35413    1.80583  -0.196    0.845
COUNTRYMEXICO    -3.18656    2.30068  -1.385    0.166
COUNTRYPERU      -4.32287    3.73073  -1.159    0.247
EUROPE          -12.98235    3.27905  -3.959 7.74e-05 ***
AFRICA            2.94095    5.93871   0.495    0.620
meanSTATURECM     0.53621    0.08606   6.231 5.45e-10 ***
nbEDUCATION       5.31460    0.67500   7.873 5.14e-15 ***
BMI               1.64558    0.14165  11.617  < 2e-16 ***
wealth_1         -0.01695    0.33702  -0.050    0.960
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25.87 on 2450 degrees of freedom
  (4475 observations deleted due to missingness)
Multiple R-squared: 0.481,      Adjusted R-squared: 0.4784
F-statistic: 189.2 on 12 and 2450 DF,  p-value: < 2.2e-16
```

PC1

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.814e-03  3.414e-03   1.117 0.264008
AGE             -3.111e-04  4.956e-05  -6.278 4.05e-10 ***
COUNTRYCHILE    -6.548e-03  1.751e-03  -3.740 0.000188 ***
COUNTRYCOLOMBIA -8.649e-04  1.496e-03  -0.578 0.563119
COUNTRYMEXICO   -1.135e-02  1.925e-03  -5.898 4.19e-09 ***
COUNTRYPERU     -4.200e-03  3.132e-03  -1.341 0.180051
EUROPE           1.838e-02  2.708e-03   6.788 1.42e-11 ***
AFRICA           3.555e-02  4.981e-03   7.138 1.24e-12 ***
nbEDUCATION     -7.779e-04  5.641e-04  -1.379 0.168033
wealth_1        -2.648e-04  2.837e-04  -0.933 0.350757
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0218 on 2454 degrees of freedom
  (4474 observations deleted due to missingness)
Multiple R-squared: 0.1519,Adjusted R-squared: 0.1488
F-statistic: 48.85 on 9 and 2454 DF,  p-value: < 2.2e-16
```

PC2

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.634e-02  2.866e-03   5.701 1.33e-08 ***
AGE             -3.588e-05  4.160e-05  -0.863 0.388391
COUNTRYCHILE    -5.466e-03  1.469e-03  -3.720 0.000203 ***
COUNTRYCOLOMBIA -5.839e-03  1.255e-03  -4.651 3.47e-06 ***
COUNTRYMEXICO   -5.116e-03  1.615e-03  -3.167 0.001559 **
COUNTRYPERU     -9.527e-03  2.629e-03  -3.624 0.000296 ***
EUROPE          -7.488e-03  2.273e-03  -3.295 0.001000 ***
AFRICA          -1.873e-02  4.180e-03  -4.480 7.80e-06 ***
nbEDUCATION     -1.023e-03  4.735e-04  -2.160 0.030840 *
wealth_1        -6.436e-04  2.381e-04  -2.703 0.006921 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01829 on 2454 degrees of freedom
  (4474 observations deleted due to missingness)
Multiple R-squared: 0.0238,Adjusted R-squared: 0.02022
F-statistic: 6.647 on 9 and 2454 DF,  p-value: 1.92e-09
```

PC3

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.598e-02  2.435e-03   6.565 6.32e-11 ***
AGE             -4.266e-04  3.534e-05 -12.070  < 2e-16 ***
COUNTRYCHILE     5.268e-03  1.248e-03   4.220 2.53e-05 ***
COUNTRYCOLOMBIA -1.785e-03  1.067e-03  -1.674   0.0943 .
COUNTRYMEXICO    1.003e-02  1.373e-03   7.310 3.59e-13 ***
COUNTRYPERU      6.095e-03  2.234e-03   2.729   0.0064 **
EUROPE          -1.388e-02  1.931e-03  -7.185 8.85e-13 ***
AFRICA           7.203e-04  3.552e-03   0.203   0.8393
nbEDUCATION      9.882e-05  4.023e-04   0.246   0.8060
wealth_1        -2.317e-04  2.023e-04  -1.146   0.2521
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01554 on 2454 degrees of freedom
  (4474 observations deleted due to missingness)
Multiple R-squared: 0.2138,Adjusted R-squared: 0.2109
F-statistic: 74.15 on 9 and 2454 DF,  p-value: < 2.2e-16
```

PC4

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -9.814e-03  2.006e-03  -4.893 1.06e-06 ***
AGE              3.903e-04  2.911e-05  13.405  < 2e-16 ***
COUNTRYCHILE     1.377e-03  1.028e-03   1.339   0.1807
COUNTRYCOLOMBIA -1.075e-03  8.787e-04  -1.223   0.2214
```

```
COUNTRYMEXICO   -6.463e-03  1.131e-03  -5.715 1.23e-08 ***
COUNTRYPERU     -2.011e-03  1.840e-03  -1.093   0.2744
EUROPE           8.914e-03  1.591e-03   5.603 2.34e-08 ***
AFRICA          -1.589e-02  2.926e-03  -5.431 6.14e-08 ***
nbEDUCATION     -8.285e-04  3.314e-04  -2.500   0.0125 *
wealth_1         6.505e-05  1.667e-04   0.390   0.6963
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0128 on 2454 degrees of freedom
  (4474 observations deleted due to missingness)
Multiple R-squared: 0.1842,Adjusted R-squared: 0.1813
F-statistic: 61.58 on 9 and 2454 DF,  p-value: < 2.2e-16
```

PC5

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1.444e-03  2.131e-03  -0.678  0.49798
AGE             -2.698e-04  3.093e-05  -8.721  < 2e-16 ***
COUNTRYCHILE     3.412e-03  1.093e-03   3.122  0.00182 **
COUNTRYCOLOMBIA -4.664e-04  9.335e-04  -0.500  0.61740
COUNTRYMEXICO    2.004e-03  1.201e-03   1.668  0.09541 .
COUNTRYPERU     -4.614e-03  1.955e-03  -2.360  0.01834 *
EUROPE           7.047e-03  1.690e-03   4.169 3.16e-05 ***
AFRICA           1.648e-03  3.109e-03   0.530  0.59599
nbEDUCATION      1.016e-03  3.521e-04   2.886  0.00394 **
wealth_1         4.609e-05  1.771e-04   0.260  0.79464
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0136 on 2454 degrees of freedom
  (4474 observations deleted due to missingness)
Multiple R-squared: 0.04816,     Adjusted R-squared: 0.04466
F-statistic: 13.79 on 9 and 2454 DF,  p-value: < 2.2e-16
```

Supplementary Note S3. Description of face shape changes associated with PCs.

PC1

The first PC explains changes on the overall shape of the face, with shorter, narrower and anteriorly projected faces occupying the positive values of the PC axis. Also, positive scores across the first PC are characterized by shorter noses, eyes displaced downwards and lips displaced upwards.

PC2

Positive values across the second PC correspond to shorter and wider faces, depicting also a more superior position of the lips, and a lower localization of the eyes. Anterior projection of the chin is characteristic of individuals occupying the positive values, whereas the nose and the glabellar region are more projected on the negative values of the second PC axis.

PC3

PC3 is mainly dominated by facial flatness, with more flattened faces on the negative values. Facial flatness can be seen as composite changes including the anterior displacement of the eyes and frontal region, along with a posterior movement of lips, chin and the inferior face in general. Negative values across this PC (flattened faces) also present more lateralized and downwardly displaced earlobes.

PC4

PC4 can be seen as a descriptor of changes occurring in the superior face, more specifically in the position of the frontotemporale. Individuals occupying the negative scores of this PC present marked glabellar and nasal projection accompanied by a downward displacement of the frontotemporal relative to the eyes. In contrast, the superior lips are elevated towards the inferior part of the nose on individuals placed at the positive values.

PC5

The fifth PC describes variations attaining the whole configuration of landmarks, and involving a downward displacement and narrowing of the nose, a shortening of the lips, and a more superior position of the superciliares in the positive values. Conversely, negative scores are depicted by more laterally placed eyes, more medially placed earlobes, and wider frontals.

Supplementary Tables

Supplementary Table S1. Allele frequencies at 30 SNP markers selected for African, European and Native American ancestry estimation and the absolute difference in reference allele frequency (Δ) between continental populations.

		CONTINENTAL POPULATION			Δ			
SNP	Allele	AFRICA	AMERICA	EUROPE	Afr-Ame	Afr-Eur	Ame-Eur	Mean Δ
rs1544450	G	0.95	0.00	0.09	0.95	0.86	0.09	0.63
rs1834619	G	0.00	0.97	0.04	0.97	0.04	0.94	0.65
rs356652	T	0.00	0.93	0.07	0.93	0.07	0.87	0.62
rs260690	C	0.64	0.96	0.05	0.32	0.60	0.92	0.61
rs2176046	G	0.01	0.93	0.06	0.92	0.05	0.87	0.61
rs10510511	G	0.00	0.92	0.02	0.92	0.02	0.89	0.61
rs3870336	G	0.09	0.94	0.09	0.85	0.00	0.85	0.57
rs10935320	T	0.15	0.98	0.10	0.83	0.05	0.88	0.58
rs11725412	A	0.21	1.00	0.06	0.79	0.15	0.94	0.63
rs10037656	A	0.34	0.98	0.10	0.64	0.24	0.88	0.59
rs4145160	G	0.09	0.91	0.07	0.82	0.03	0.84	0.56
rs1559163	A	0.00	0.85	0.02	0.85	0.02	0.83	0.57
rs2042314	C	0.14	1.00	0.15	0.86	0.01	0.85	0.57
rs12662498	G	0.01	0.98	0.07	0.97	0.06	0.91	0.65
rs17086231	C	0.02	0.94	0.12	0.93	0.10	0.83	0.62
rs6464749	A	0.89	0.00	0.05	0.89	0.84	0.05	0.60
rs7018273	A	0.86	0.00	0.02	0.86	0.84	0.02	0.57
rs12347078	A	0.88	0.00	0.04	0.88	0.84	0.04	0.58
rs734241	G	0.04	0.99	0.07	0.94	0.02	0.92	0.63
rs174570	C	0.01	1.00	0.11	0.99	0.10	0.89	0.66
rs7134749	T	0.21	0.90	0.03	0.69	0.18	0.87	0.58
rs2052386	G	0.08	0.93	0.10	0.85	0.02	0.83	0.57
rs1849384	A	0.97	0.00	0.08	0.97	0.89	0.08	0.65
rs4769128	C	0.15	0.99	0.13	0.83	0.02	0.86	0.57
rs1243370	T	0.24	0.92	0.06	0.68	0.18	0.86	0.58
rs2719921	G	0.88	0.00	0.03	0.88	0.84	0.03	0.58
rs1197062	T	0.89	0.00	0.06	0.89	0.83	0.06	0.59
rs12104228	T	0.87	0.00	0.01	0.87	0.86	0.01	0.58
rs6119879	C	0.69	0.00	0.84	0.69	0.15	0.84	0.56
rs2426552	C	0.83	0.00	0.01	0.83	0.83	0.01	0.56

Supplementary Table S2: Population samples used for genetic ancestry estimation.

AFRICA	N
YRI	109
Mandenka	22
Yoruba	21
BantuKenya	9
BantuSouthAfrica	8
Total	169

EUROPE	N
CEU	111
TSI	88
French	28
Sardinian	28
Basque	24
Italian	12
Tuscan	8
Total	299

AMERICA	N
PimaMX	33
Cabecar	31
Surui	24
Aymara	23
Zapotec	23
PimaAZ	22
Quechua2	22
Maya	18
Quechua1	18
Mixe	17
Kaqchikel	13
Karitiana	13
Ticuna2	12
Wayuu	12
Inga	10
Chilote	8
Piapoco	7
Guahibo	6
Guarani	6
Ticuna1	6
Arhuaco	5
Diaguita	5
Embera	5
Guaymi	5
Mixtec	5
Wichi	5
Zenu	5
Bribri	4
Chono	4
Hulliche	4
Kogi	4
Toba	4
Yaghan	4
Maleku	3
Palikur	3
Teribe	3
Waunana	3
Chane	2
Kaingang	2
Kalina	2
Arara	1
Chorotega	1
Huetar	1
Jamamadi	1
Parakana	1
Purepecha	1
Yaqui	1
Total	408

Supplementary Table S3: Coefficients and P-values for regression of pair-wise inter-landmark distances and ancestry.

Note: Only data for P-values<10^{-6} with European ancestry (133/427) are shown (ranked by European P-value). Landmarks are labelled as in Supplementary Figure 1. Equivalent distances between the two sides of the face have been averaged.

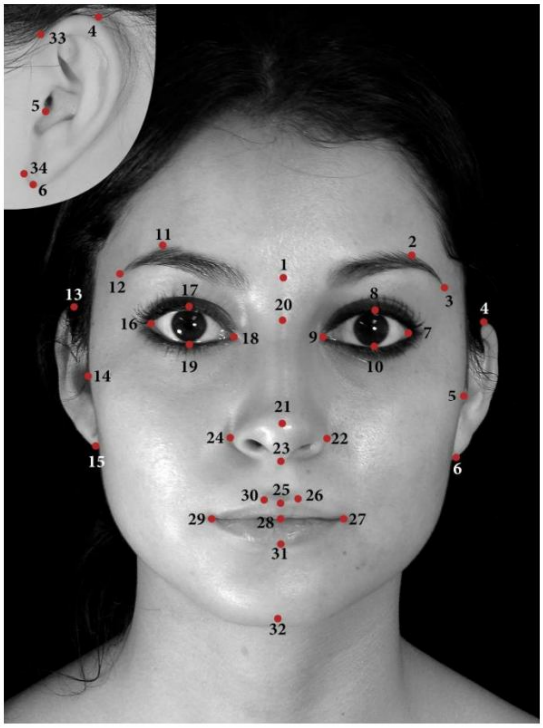
		African ancestry		European Ancestry	
Distance #	Landmarks	Coef.	P-value	Coef.	P-value
1	28.26.30	0.9	1.27E-02	-1.9	2.00E-16
2	28.22.24	2.2	1.18E-03	-3.1	2.58E-16
3	31.26.30	2.9	3.07E-05	-3.2	3.13E-16
4	31.22.24	3.8	3.44E-05	-4.0	2.34E-15
5	22.29.24.27	2.5	4.34E-03	-3.8	4.49E-15
6	27.24.29.22	2.5	4.34E-03	-3.8	4.49E-15
7	5.34.14.36	-3.5	2.65E-07	-2.9	1.11E-14
8	22.24	3.5	5.93E-06	-3.2	1.10E-13
9	26.29.30.27	1.2	9.82E-02	-3.0	1.59E-13
10	27.30.29.26	1.2	9.82E-02	-3.0	1.59E-13
11	31.2.11	0.6	7.31E-01	-7.5	6.32E-13
12	25.28	1.0	4.80E-03	-1.4	4.27E-12
13	25.31	3.1	1.08E-05	-2.7	6.38E-12
14	25.27.29	1.3	4.86E-02	-2.4	7.69E-12
15	28.2.11	-1.1	5.22E-01	-6.5	7.78E-12
16	5.36.14.34	-4.6	4.31E-02	-8.5	8.89E-12
17	34.14.36.5	-4.6	4.31E-02	-8.5	8.89E-12
18	32.26.30	4.0	5.55E-04	-4.3	9.81E-12
19	2.29.11.27	-0.8	6.56E-01	-6.8	9.89E-12
20	27.11.29.2	-0.8	6.56E-01	-6.8	9.89E-12
21	31.3.12	0.6	7.31E-01	-6.9	1.31E-11
22	34.36	-3.1	1.73E-01	-8.4	1.60E-11
23	23.22.24	1.3	4.73E-03	-1.7	1.73E-11
24	3.29.12.27	-1.2	4.90E-01	-6.7	1.88E-11
25	27.12.29.3	-1.2	4.90E-01	-6.7	1.88E-11
26	31.10.19	0.3	8.50E-01	-5.2	2.08E-11
27	10.29.19.27	-0.6	6.33E-01	-4.8	2.96E-11
28	27.19.29.10	-0.6	6.33E-01	-4.8	2.96E-11
29	22.30.24.26	2.1	2.16E-03	-2.5	5.31E-11
30	26.24.30.22	2.1	2.16E-03	-2.5	5.31E-11
31	25.22.24	1.9	4.22E-03	-2.4	5.46E-11
32	27.29	1.0	2.66E-01	-3.3	9.35E-11
33	26.27.30.29	1.3	3.71E-02	-2.1	1.27E-10
34	31.7.16	1.2	4.42E-01	-5.4	1.45E-10
35	3.36.12.34	-3.7	1.13E-01	-8.1	2.00E-10
36	34.12.36.3	-3.7	1.13E-01	-8.1	2.00E-10
37	2.27.11.29	-1.6	3.29E-01	-5.8	2.01E-10
38	28.3.12	-0.9	5.98E-01	-6.0	2.02E-10
39	31.27.29	1.4	2.62E-02	-2.2	2.17E-10
40	2.24.11.22	-0.4	7.98E-01	-5.1	2.26E-10

41	22.11.24.2	-0.4	7.98E-01	-5.1	2.26E-10
42	33.36.35.34	-6.4	5.97E-03	-8.2	2.43E-10
43	34.35.36.33	-6.4	5.97E-03	-8.2	2.43E-10
44	28.10.19	-1.3	2.83E-01	-4.2	3.18E-10
45	25.32	4.2	2.30E-04	-3.9	3.53E-10
46	7.29.16.27	0.1	9.70E-01	-5.0	5.81E-10
47	27.16.29.7	0.1	9.70E-01	-5.0	5.81E-10
48	20.31	0.2	8.90E-01	-5.1	6.17E-10
49	23.31	2.9	1.13E-03	-3.0	6.61E-10
50	3.27.12.29	-1.4	3.68E-01	-5.5	7.65E-10
51	5.14	-5.1	2.69E-02	-7.8	8.80E-10
52	2.8.11.17	0.6	4.17E-01	-2.4	8.93E-10
53	28.27.29	0.9	1.44E-01	-1.9	1.43E-09
54	32.2.11	0.4	8.47E-01	-7.5	1.66E-09
55	9.29.18.27	-1.0	4.13E-01	-4.0	2.06E-09
56	27.18.29.9	-1.0	4.13E-01	-4.0	2.06E-09
57	6.36.15.34	-2.0	3.94E-01	-7.6	2.28E-09
58	34.15.36.6	-2.0	3.94E-01	-7.6	2.28E-09
59	28.7.16	-0.3	8.45E-01	-4.4	2.31E-09
60	10.27.19.29	-2.5	2.80E-02	-3.7	2.42E-09
61	31.8.17	0.0	9.76E-01	-5.0	2.89E-09
62	5.15.14.6	-3.4	1.44E-01	-7.6	3.29E-09
63	6.14.15.5	-3.4	1.44E-01	-7.6	3.29E-09
64	5.6.14.15	-2.3	2.88E-04	-2.1	3.36E-09
65	32.3.12	-0.1	9.59E-01	-7.1	4.43E-09
66	8.29.17.27	-1.0	4.58E-01	-4.6	4.59E-09
67	27.17.29.8	-1.0	4.58E-01	-4.6	4.59E-09
68	4.36.13.34	-5.7	2.15E-02	-8.0	4.95E-09
69	34.13.36.4	-5.7	2.15E-02	-8.0	4.95E-09
70	31.9.18	0.7	5.91E-01	-4.4	5.13E-09
71	10.24.19.22	0.8	4.35E-01	-3.2	5.66E-09
72	22.19.24.10	0.8	4.35E-01	-3.2	5.66E-09
73	3.34.12.36	-1.9	2.75E-01	-5.5	6.27E-09
74	25.2.11	-1.9	2.59E-01	-5.2	1.53E-08
75	23.27.29	0.5	5.01E-01	-2.1	1.94E-08
76	23.2.11	-2.1	1.53E-01	-4.5	2.24E-08
77	20.28	-1.8	1.66E-01	-4.0	2.26E-08
78	25.26.30	0.1	7.34E-01	-0.8	2.51E-08
79	3.24.12.22	-0.8	6.16E-01	-4.6	2.91E-08
80	22.12.24.3	-0.8	6.16E-01	-4.6	2.91E-08
81	33.34.35.36	-5.5	1.11E-07	-3.1	3.13E-08
82	5.35.14.33	-6.3	7.54E-03	-7.1	3.36E-08
83	33.14.35.5	-6.3	7.54E-03	-7.1	3.36E-08
84	7.27.16.29	-1.4	2.54E-01	-3.8	3.44E-08
85	1.31	1.4	4.18E-01	-5.2	4.20E-08
86	2.30.11.26	-1.7	3.12E-01	-5.1	4.68E-08
87	26.11.30.2	-1.7	3.12E-01	-5.1	4.68E-08
88	28.8.17	-1.8	1.87E-01	-4.0	5.98E-08

89	32.10.19	-0.3	8.73E-01	-5.3	6.58E-08
90	6.35.15.33	-5.1	3.33E-02	-7.1	6.79E-08
91	33.15.35.6	-5.1	3.33E-02	-7.1	6.79E-08
92	6.34.15.36	2.8	4.44E-06	1.8	7.73E-08
93	2.9.11.18	0.3	7.48E-01	-2.3	9.34E-08
94	3.14.12.5	-2.9	2.02E-01	-6.7	9.62E-08
95	5.12.14.3	-2.9	2.02E-01	-6.7	9.62E-08
96	28.9.18	-1.1	3.70E-01	-3.4	1.01E-07
97	20.27.29	-1.8	1.74E-01	-3.8	1.02E-07
98	2.18.11.9	0.3	8.14E-01	-3.5	1.16E-07
99	9.11.18.2	0.3	8.14E-01	-3.5	1.16E-07
100	22.27.24.29	0.7	2.71E-01	-1.9	1.46E-07
101	6.15	-1.1	6.47E-01	-6.9	1.58E-07
102	32.22.24	3.6	8.78E-03	-3.9	1.63E-07
103	2.22.11.24	-3.0	1.50E-02	-3.5	1.72E-07
104	32.7.16	0.4	8.46E-01	-5.3	2.41E-07
105	7.24.16.22	1.2	3.15E-01	-3.3	2.73E-07
106	22.16.24.7	1.2	3.15E-01	-3.3	2.73E-07
107	5.29.14.27	1.9	3.61E-01	-5.8	2.98E-07
108	27.14.29.5	1.9	3.61E-01	-5.8	2.98E-07
109	8.24.17.22	-0.1	9.09E-01	-3.0	3.00E-07
110	22.17.24.8	-0.1	9.09E-01	-3.0	3.00E-07
111	23.28	0.7	2.55E-01	-1.7	3.50E-07
112	23.32	3.3	1.48E-02	-3.8	3.61E-07
113	8.27.17.29	-2.5	5.29E-02	-3.6	3.81E-07
114	9.27.18.29	-1.5	1.99E-01	-3.2	3.82E-07
115	7.36.16.34	-0.7	7.46E-01	-5.8	3.99E-07
116	34.16.36.7	-0.7	7.46E-01	-5.8	3.99E-07
117	25.3.12	-1.4	3.93E-01	-4.7	4.11E-07
118	9.36.18.34	-0.9	6.38E-01	-5.2	4.46E-07
119	34.18.36.9	-0.9	6.38E-01	-5.2	4.46E-07
120	2.26.11.30	-2.0	2.26E-01	-4.5	4.70E-07
121	4.15.13.6	-4.5	7.71E-02	-7.1	5.27E-07
122	6.13.15.4	-4.5	7.71E-02	-7.1	5.27E-07
123	10.36.19.34	0.0	9.92E-01	-5.6	5.63E-07
124	34.19.36.10	0.0	9.92E-01	-5.6	5.63E-07
125	3.30.12.26	-1.4	4.02E-01	-4.6	5.96E-07
126	26.12.30.3	-1.4	4.02E-01	-4.6	5.96E-07
127	4.14.13.5	-5.5	2.46E-02	-6.8	5.99E-07
128	5.13.14.4	-5.5	2.46E-02	-6.8	5.99E-07
129	20.32	0.2	9.14E-01	-5.3	7.63E-07
130	32.8.17	-0.2	8.98E-01	-5.2	8.08E-07
131	4.34.13.36	-4.6	8.47E-05	-3.2	8.35E-07
132	3.15.12.6	-1.3	5.78E-01	-6.5	9.32E-07
133	6.12.15.3	-1.3	5.78E-01	-6.5	9.32E-07

Supplementary Figures

Supplementary Figure S1: Position of facial landmarks.



Landmarks:

1. Glabella
2. Superciliare (left)
3. Frontotemporale (left)
4. Superaurale (left)
5. Superior lateral tragion (left)
6. Subaurale (left)
7. Exocanthion (left)
8. Palpebrale superiorus (left)
9. Endocanthion (left)
10. Palpebrale inferiorus (left)
11. Superciliare (right)
12. Frontotemporale (right)
13. Superaurale (right)
14. Superior lateral tragion (right)
15. Subaurale (right)
16. Exocanthion (right)
17. Palpebrale superiorus (right)
18. Endocanthion (right)
19. Palpebrale inferiorus (right)
20. Sellion
21. Infrapronasale
22. Alare (left)
23. Subnasale
24. Alare (right)
25. Labiale superiorus (sagital)
26. Labiale superiorus (left)
27. Cheilion (left)
28. Stomion
29. Cheilion (right)
30. Labiale superiorus (right)
31. Labiale inferiorus
32. Pogonion
33. Otobasion superiorious (left)
34. Otobasion inferiorous (left)
35. Otobasion superiorious (right)
36. Otobasion inferiorous (right)

Supplementary Figure S2: Facial morphs displaying the shape changes associated with PCs 1 to 5.