UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CURSO DE BIOMEDICINA
ESTÁGIO DE PESQUISA E TRABALHO DE CONCLUSÃO DE CURSO EM BIOMEDICINA

CÁSSIO AUGUSTO RODRIGUES BETTIM

MC1R E HETEROCLASSIFICAÇÃO DE FENÓTIPOS DE FACE NA POPULAÇÃO DO RIO GRANDE DO SUL

Porto Alegre
2022

CÁSSIO AUGUSTO RODRIGUES BETTIM

MC1R E HETEROCLASSIFICAÇÃO DE FENÓTIPOS DE FACE NA POPULAÇÃO DO RIO GRANDE DO SUL

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel em Biomedicina.

Orientador: Dr. Eduardo Filipe Ávila Silva
Coorientador: Alexsandro Vasconcellos da Silva

CIP - Catalogação na Publicação

Bettim, Cássio Augusto Rodrigues
MC1R e heteroclassificação de fenótipos de face na população do Rio Grande do Sul / Cássio Augusto Rodrigues Bettim. -- 2022.
76 f.
Orientador: Eduardo Filipe Avila.

Coorientador: Alexsandro Vasconcellos.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciências Básicas da Saúde, Curso de Biomedicina, Porto Alegre, BR-RS, 2022.

1. Fenotipagem forense através do DNA. 2. Características externamente visíveis. 3. Idade aparente. 4. Gene MC1R. 5. Polimorfismos de nucleotídeo único. I. Avila, Eduardo Filipe, orient. II. Vasconcellos, Alexsandro, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

CÁSSIO AUGUSTO RODRIGUES BETTIM

MC1R E HETEROCLASSIFICAÇÃO DE FENÓTIPOS DE FACE NA POPULAÇÃO DO RIO GRANDE DO SUL

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel em Biomedicina.

Aprovado em 22 de Junho de 2022

BANCA EXAMINADORA

_____

Drª. Catieli Gobetti Lindholz - PUCRS

_____

Dr. Gustavo Chemale - Perito Criminal Federal

_____

Dr. Eduardo Filipe Ávila Silva - PUCRS

_____

Alexsandro Vasconcellos da Silva - PUCRS

AGRADECIMENTOS

Agradeço a meus pais, Cintia Bettim e Alisson Bettim, pelo apoio incondicional desde sempre, pelo suporte proporcionado para que eu pudesse me aventurar longe de casa e buscar a realização de meus sonhos, e por sempre me estimularem a dar o meu melhor e acreditar em mim mesmo. Toda conquista minha também é de vocês. Amo-lhes infinitamente e lhes sou eternamente grato.

Agradeço a meu orientador, Eduardo Avila, a Profª Clarice Alho, e a todos os membros e pesquisadores incríveis do Laboratório de Genética Forense da PUCRS - em especial, "Camilinha", "Cati", Lucas, Alessandro, Pietro e "Vasco" - por terem me acolhido como aluno de iniciação científica, pela paciência e pelos ensinamentos durante estes últimos anos e na realização deste trabalho de conclusão, e claro, pelos momentos divertidos, de alegria contagiante e festas que só vocês sabem organizar. Mais do que colegas de laboratório, hoje os considero grandes amigos.

Agradeço a meus colegas, amigos, companheiros de jornada acadêmica e de vida, Arthur, Ana, Daphne, Giovanna, Geciele, "Gabão", Letícia, Luiza, "Malu", Mailton, Vitória e Wesley, que fizeram deste período peculiar da graduação tão especial. Rimos juntos, choramos juntos, comemoramos as conquistas uns dos outros, compartilhamos histórias e momentos memoráveis. Graças a vocês, a rotina, mesmo cansativa e até exaustiva em alguns momentos, no fim das contas, era leve e prazerosa.

Por fim, agradeço a meus amigos de longa data, praticamente irmãos, Giordano e Wagner, pelo companheirismo com mais de 12 anos de história, que nem o tempo e a distância foram capazes de enfraquecer. Apesar dos caminhos diferentes que tomamos na vida, sempre estivemos ali um pelo outro, nos apoiando. Obrigado por fazerem parte da minha história.

RESUMO

A Fenotipagem forense através do DNA consiste na utilização de metodologias para predição de características externamente visíveis (CEVs) a partir do material genético de amostras biológicas encontradas em locais de crimes, e tem se mostrado uma ferramenta promissora no auxílio da identificação humana em atividades policiais. Atualmente, métodos baseados em ensaios multiplex e modelos estatísticos de predição de CEVs relacionadas à pigmentação de cabelo, pele e íris utilizando painéis de biomarcadores do tipo SNP e Indel já foram desenvolvidos e validados pela comunidade científica forense. Assim como traços de pigmentação, a idade aparente (*PA*, do inglês *perceived age*) de um indivíduo também pode ser considerada uma CEV e sua estimação em indivíduos desconhecidos pode ser útil para otimização no andamento de investigações. Liu et al. (2016) foi pioneiro em evidenciar que, além do estilo de vida e fatores ambientais, a presença de variantes genéticas do tipo SNP e Indel no gene MC1R - que codifica um receptor transmembrana responsável pela regulação da produção de melanina - e locus adjacentes também parece contribuir para a PA de um indivíduo. O grupo destacou a associação deste painel de polimorfismos e a idade aparente na população europeia, onde portadores de haplótipos de risco aparentavam ser até 2 anos mais velhos do que sua idade cronológica (*CA*, do inglês *Chronological Age*). Entendendo que relações genótipo-fenótipo não podem ser extrapoladas entre diferentes grupos populacionais, o objetivo principal deste estudo foi testar essa hipótese e verificar a presença da mesma associação na população miscigenada do Rio Grande do Sul. A partir dados genômicos de uma amostra de 261 indivíduos representativos da população gaúcha e utilizando um modelo de regressão linear múltipla, nosso grupo foi capaz de verificar uma associação significativa entre 9 variantes intrônicas em locus adjacentes ao MC1R (*e.g.:* AFG3L1P, TUBB3, FANCA, etc.) e aparência etária facial, cuja PA foi definida após heteroclassificação etária de imagens frontais de face através de 11 avaliadores. Curiosamente, diferente do constatado em populações europeias, nossos resultados mostram que a presença dos alelos de efeito (R) das variantes em nossa amostra influenciou tanto em fenótipos de face mais jovens quanto em fenótipos de face mais velhos. A influência de cada variante na PA é representada pelos valores β. A título de exemplo, indivíduos homozigotos para AFG3L1P rs112220510 aparentam ser 3,16±1,34 anos mais jovens do que não portadores da variante, enquanto homozigotos para AFG3L1P rs201156703 parecem ser 7,54±3,59 anos mais velhos do que não portadores. Apesar de apresentarmos propostas sobre os mecanismos moleculares por detrás da influência destes polimorfismos na PA, estudos futuros ainda serão necessários para que estes sejam devidamente elucidados.

Palavras-chave: Fenotipagem forense através do DNA; Características externamente visíveis; Idade aparente; MC1R; Polimorfismos de nucleotídeo único.

ABSTRACT

Forensic DNA Phenotyping (FDP) consists of the use of methodologies for predicting externally visible characteristics (EVCs) from the genetic material of biological samples found in crime scenes, and has proven to be a promising tool in aiding human identification in police activities. Currently, methods based on multiplex assays and statistical models of prediction of EVCs related to hair, skin and iris pigmentation using panels of SNP and INDEL biomarkers have already been developed and validated by the forensic scientific community. As well as traces of pigmentation, an individual's perceived age (PA) can also be considered an EVC and its estimation in unknown individuals can be useful for optimization in the progress of investigations. Liu et al. (2016) were pioneers in evidencing that, in addition to lifestyle and environmental factors, the presence of SNP and Indel variants in the MC1R gene - which encodes a transmembrane receptor responsible for regulating melanin production - and adjacent loci also seems to contribute to an individual's PA. The group highlighted the association between these MC1R gene polymorphisms and the PA in the European population, where carriers of risk haplotypes appeared to be up to 2 years older in comparison to their chronological age (CA). Understanding that genotype-phenotype relationships cannot be extrapolated between different population groups, the main objective of this study was to test this hypothesis and verify the applicability of this variant panel in the Rio Grande do Sul admixed population. Based on genomic data from a sample of 261 individuals representative of gaucho population and using a multiple linear regression (MLR) model, our group was able to verify a significant association between 9 intronic variants in locus adjacent to MC1R (*e.g.*: AFG3L1P, TUBB3, FANCA, etc.) and facial age appearance, whose PA was defined after age heteroclassification of frontal face images through 11 assessors. Interestingly, different from that observed in European populations, our results show that the presence of effect alleles (R) of the selected variants in our sample influenced both younger and older face phenotypes. The influence of each variant on PA is expressed as β values. For example, homozygotes for AFG3L1P rs112220510 appear to be 3.16±1.34 years younger than non-carriers of the variant, while homozygotes for AFG3L1P rs201156703 appear to be 7.54±3.59 years older than non-carriers. Although we present hypotheses about the molecular mechanisms behind the influence of these polymorphisms on PA in this study, future studies will still be necessary for them to be properly elucidated.

Keywords: Forensic DNA Phenotyping; External Visible Characteristics; Perceived Age; MC1R; Single Nucleotide Polymorphisms.

LISTA DE ABREVIATURAS

AIM - Marcadores informativos de ancestralidade (*Ancestry Informative Markers*)

CODIS - Sistema combinado de índices de ADN (*Combined DNA Index System*)

DNA - Ácido desoxirribonucleico (*Deoxyribonucleic Acid*)

EVC - Características externamente visíveis (*External Visible Characteristics*)

FDP - Fenotipagem Forense através do DNA (*Forensic DNA Phenotyping*)

GWAS - Estudo de associação genômica amplo (*Genomic Wide Association Studies*)

Indel - Inserção/Deleção (*Insertion/Deletion*)

STR - Repetições curtas em tandem (*Short Tandem Repeats*)

***MC1R* -** Receptor de melanocortina tipo 1 (*Melanocortin 1 Receptor*)

mRNA - Ácido ribonucleico mensageiro (*Messenger Ribonucleic Acid*)

mtDNA - ADN mitocondrial (*mitocrondrial DNA*)

PCR - Reação em cadeia da polimerase (*Polimerase Chain Reaction*)

PI - Probabilidade de identidade (*Probability of Identity)*

sjTREC - Círculo de excisão do receptor de células T (*Signal Joint T-Cell Receptor Excision Circle*)

SNP - Polimorfismo de nucleotídeo único (*Single Nucleotide Polimorfism*)

UV - Ultravioleta (*Ultraviolet*)

SUMÁRIO

1 INTRODUÇÃO

1.1 O CONTEXTO ATUAL DA GENÉTICA FORENSE E DA IDENTIFICAÇÃO HUMANA

Nas últimas décadas, a genética forense tornou-se uma ferramenta indispensável no que se refere à identificação humana e resolução de problemas jurídicos, graças ao esforço conjunto de estudiosos nas áreas da genética humana, médica e de populações, associados aos avanços nas técnicas de biologia molecular. Evidências genéticas podem ser obtidas de qualquer vestígio biológico humano - sangue, saliva, sêmen, cabelo, ossos, etc - e são peças-chave para elucidação de casos onde se faz necessária a caracterização de indivíduos desconhecidos em cenas de crime, de pessoas desaparecidas, de vítimas de desastres naturais ou para confirmação de parentesco.

Na genética forense contemporânea, as *short tandem repeats* (*STRs*) - também denominadas de sequências microssatélites - autossômicas são marcadores moleculares amplamente utilizados e consideradas o padrão ouro no que diz respeito à identificação humana (KOWALCZYK *et al.*, 2018). As *STRs* são uma classe de pequenas sequências de DNA repetidas em tandem, que variam de 2 a 6 nucleotídeos e representam cerca de 3% do genoma humano (LANDER *et al.*, 2001). Tratam-se de marcadores de natureza polimórfica, podendo variar tanto no número de unidade de repetições quanto na rigorosidade do padrão de repetição, tendo caráter multialélico e apresentando alta variabilidade entre indivíduos - o que lhes confere alto poder de discriminação quando utilizados para propósitos de identificação (DIAS FILHO *et al.*, 2020; WYNER; BARASH; MCNEVIN, 2020). Associar a variabilidade destes marcadores à análise de múltiplos *loci* também reduz a *probability of identity (PI)* - probabilidade de que dois indivíduos selecionados ao acaso apresentem o mesmo genótipo para um dado *locus*. Por conta disso, geneticistas forenses optam por utilizar kits comerciais que permitem a amplificação simultânea de 15, 16, 18, 21 e até 27 *loci STR* com características únicas em técnicas empregando eletroforese capilar e *Multiplex Polymerase Chain Reaction (PCR)*, cujo *PI value* alcança valores excepcionais de até 9.09×10^{-31} (POWERPLEX® FUSION 6C SYSTEM, [*s. d.*]).

O número de *loci STR* amplificados em uma análise padrão tende a aumentar à medida que os bancos de dados de perfis genéticos também aumentam. Essa tendência é esperada, visto que a utilização destes marcadores para fins de caracterização individual depende da comparação direta com perfis genéticos de referência - seja de um possível suspeito levantado durante investigação, de familiares de primeiro grau deste ou confronto direto com perfis já depositados nestes bancos de dados genéticos (SCHNEIDER; PRAINSACK; KAYSER,

2019).Como exemplo nacional, podemos citar a rede integrada de bancos de perfis genéticos (RIBPG), iniciada em 2009 e instituída pelo decreto nº 7950/2013 e que hoje conta com mais de 100 mil perfis *STR* cadastrados no sistema CODIS (*Combined DNA Index System*) - implementado a partir de uma parceria entre Polícia Federal e FBI (*Federal Bureau of Investigation*) -, compostos majoritariamente de amostras de condenados, seguido de vestígios de locais de crime, restos mortais não identificados, e de familiares de pessoas desaparecidas. Segundo o último relatório semestral disponível, a RIBPG conta, atualmente, com o suporte de 20 laboratórios estaduais - incluindo o Instituto Geral de Perícias do Rio Grande do Sul (IGP-RS) -, 1 laboratório distrital e 1 laboratório da polícia federal e já auxiliou em mais de 2800 investigações. Além disso, a lei 12.654, sancionada em 2012, regulamenta a coleta e manutenção de perfis genéticos de investigados e relaciona os crimes ou situações determinantes para que estes sejam armazenados nos bancos de dados (BRITO; PONTES, 2020).

Contudo, é importante salientar que, apesar de todos os avanços nas técnicas de biologia molecular e, consequentemente, na utilização dos perfis *STR* para identificação humana, existem cenários que podem impor limitações ao curso da investigação. Casos onde não é possível encontrar associação entre a evidência genética e as amostras de referência - seja de um possível suspeito ou de qualquer indivíduo presente nos bancos de dados - estão presentes na rotina do perito criminal (KAYSER, 2015). Também, por se tratar de um campo relativamente recente no campo das ciências forenses, existem divergências éticas e quanto à constitucionalidade de leis tais como a 12.654/2012 no âmbito jurídico. Por conta disso, esforços vêm sendo direcionados para o desenvolvimento de usos alternativos para amostras de DNA oriundas de locais de crime, tais como a fenotipagem forense através do DNA, ou *FDP* (do inglês, *forensic DNA phenotyping*), assunto que será abordado posteriormente neste trabalho.

1.2 POLIMORFISMOS DE NUCLEOTÍDEO ÚNICO (SNP)

Os *SNPs* consistem em *loci* onde ocorreu a substituição de uma base individual por outro nucleotídeo em uma posição específica do genoma, ocorrendo, em média, uma vez a cada 1000 pares de base e normalmente apresentando apenas dois alelos (NUSSBAUM; MCINNES; WILLARD, 2008).

Existem quatro classes principais de *SNPs* relevantes dentro da genética forense: 1) *SNPs* informativos de linhagem - encontrados no mtDNA (DNA mitocondrial) e cromossomo Y, utilizados principalmente para verificação de parentesco em casos de desaparecimento ou

desastres em massa, por conta de sua baixa taxa de mutação e ausência de recombinação; 2) *SNPs* de identificação - cumprem o mesmo propósito das *STRs*, possibilitando a individualização ou exclusão de pessoas que não poderiam ser fonte de uma amostra; 3) *SNPs* informativos de ancestralidade - também encontrados no mtDNA e cromossomo Y, e denominados *ancestry informative markers (AIMs)* -, cujos haplótipos podem inferir a ancestralidade biogeográfica por estarem presentes em diferentes frequências entre populações; 4) *SNPs* informativos de fenótipo - utilizados na inferência de características físicas por apresentam forte correlação com um fenótipo específico (BUDOWLE; VAN DAAL, 2008).

A natureza bialélica dos *SNPs* faz com que tenham menor poder discriminatório para identificação, e sendo assim são, quando necessário, utilizados como forma de complementar os resultados de análise dos perfis *STRs* (DIAS FILHO *et al.*, 2020). Estima-se, que seja necessária a análise de um conjunto de 40-60 *SNPs* para igualar-se à capacidade discriminatória de amostras de um painel de 15 *loci STR* (KOWALCZYK *et al.*, 2018). As principais vantagens dos *SNPs* em relação a outros marcadores utilizados para identificação humana são: a) potencializam resultados oriundos de amostras severamente degradadas pelo tamanho reduzido de seus produtos de *PCR*; b) taxas de mutação até cem mil vezes menores que *STRs* (10^{-8} vs. 10^{-3}); c) técnicas de sequenciamento de próxima geração possibilitam a análise massiva de um amplo quadro de *SNPs* simultâneamente, diminuindo esta diferença de poder de discriminação; d) não apresentam artefatos técnicos conhecidos como *stutter* - presentes na amplificação de *STRs* e derivados de um erro inato da enzima polimerase -, que dificultam a interpretação de amostras com misturas de perfis genéticos (BUTLER; COBLE; VALLONE, 2007).

1.3 INDELS/DIPs

São polimorfismos de comprimento definidos pela inserção ou deleção, como o próprio nome já diz, que podem variar de um a 10.000 pares de base (MULLANEY *et al.*, 2010; NUSSBAUM; MCINNES; WILLARD, 2008). Podem ser divididos em dois grupos: os multialélicos, onde se encaixam as repetições em *tandem* tais como os microssatélites, já abordados anteriormente, e minissatélites; e os bialélicos (WEBER *et al.*, 2002), cujo uso também tem sido explorado na comunidade forense .

Após identificar e caracterizar mais de 2000 polimorfismos, Weber et al. constataram propriedades básicas das *InDels*, verificando que 71% possuem variação de 2 a 4 nucleotídeos e que os bialélicos correspondem a cerca de 8% dos polimorfismos do genoma humano. Além

da ampla presença, *InDels* bialélicos também apresentam características desejáveis muito similares às dos *SNPs* - baixa taxa de mutação, produtos de amplificação pequenos e sem observação de artefatos *stutter* - e mostram uma significativa diferença de frequência entre populações (DIAS FILHO *et al.*, 2020), tornado-os marcadores informativos de ancestralidade e identificação humana, mas também úteis na predição de fenótipos, estando presentes em painéis conhecidos para tal prática (BRESLIN *et al.*, 2019).

1.4 FENOTIPAGEM FORENSE ATRAVÉS DO DNA

A *FDP* consiste na predição de características externamente visíveis (*EVCs*, do inglês *External Visible Characteristics),* como morfologia facial, altura, estimativa de idade, estimativa de idade e pigmentação de olhos, cabelo e pele (VIRMOND *et al.*, 2016) através da análise de material genético obtido de amostras biológicas, auxiliando na identificação de indivíduos desconhecidos. A *FDP* torna-se uma estratégia alternativa em situações onde: (1) o perfil genético de indivíduos suspeitos ou vítimas encontradas não corresponde à nenhum outro presente nos bancos de dados de DNA disponíveis, ou (2) quando não é possível obter amostras biológicas *ante mortem* ou de possíveis parentes próximos, em casos de desaparecidos (KAYSER, 2015). Nestes contextos, onde qualquer informação adicional acerca do doador de uma amostra é útil, a *FDP* serve como uma "testemunha biológica" e permite uma investigação policial mais direcionada.

Grande parte das *EVCs* são consideradas traços complexos, ou seja, são influenciados pela ação de múltiplos genes - onde cada gene contribui individualmente em pequena proporção para o traço fenotípico em questão - e também por fatores ambientais (KAYSER, 2015; KAYSER; SCHNEIDER, 2009). Em outras palavras, significa que estabelecer uma associação estatisticamente significativa requer o uso de milhares de amostras e a análise de milhares de marcadores *SNPs* em *genome-wide association studies (GWAS)*, tarefa inacessível para um único laboratório mas possível através de grandes esforços colaborativos, como a formação de consórcios internacionais (STRANGER; STAHL; RAJ, 2011).

Os principais exemplos da aplicação prática da *FDP* que temos atualmente consistem na predição de *EVCs* relacionados a pigmentação, visto que estes aparentam ser traços geneticamente menos complexos - apresentam alta hereditariedade e poucos genes contribuem com a maior parte do fenótipo (VIRMOND *et al.*, 2016). Métodos para predição de pigmentação da íris, cabelos e pele que utilizam marcadores *SNP* e *INDEL* já foram desenvolvidos e validados, e estas informações acerca do indivíduo já podem ser obtidas de

amostras de DNA com um poder preditivo adequado (SCHNEIDER; PRAINSACK; KAYSER, 2019).

Como um dos trabalhos pioneiros utilizando da tecnologia *FDP*, podemos citar o sistema IrisPlex desenvolvido por Walsh et al. em 2011, que consiste num painel de 6 marcadores *SNP* - HERC2 rs12913832, OCA2 rs1800407, LOC105370627 rs12896399, SLC45A2 rs16891982, TYR rs1393350 e IRF4 rs12203592 - junto de um modelo estatístico de predição, capaz de prever corretamente a coloração de íris azul ou castanha a partir de amostras genéticas com precisão acima de 90% (WALSH *et al.*, 2011). Em 2013, mais 18 novos marcadores foram adicionados ao painel, compondo o novo sistema HIrisplex, composto por 24 marcadores (23 *SNPs* e 1 *InDel*) e dois modelos de predição para cor de olhos e, também, cabelos - com acurácia de 69,5% para loiro, 78,5% para castanho, 80% para ruivo e 87,5% para preto (WALSH *et al.*, 2013). A tecnologia HIrisplex continua sendo aprimorada e, ainda em 2018, através de um artigo publicado na *Forensic Science International: Genetics*, foi introduzido a esse sistema um novo modelo de predição para coloração de pele, que passou a ser chamada de HIrisplex-S ("S" de *skin*, "pele" em inglês). O painel do HIrisplex-S conta com 41 marcadores, incluindo os de seus antecessores, de 16 genes associados à pigmentação, e pode predizer cinco categorias de cor de pele baseadas na escala de Fitzpatrick com diferentes valores de acurácia de predição - expressos pela área sob a curva característica de operação do receptor (AUC) -, sendo de 0.75 para branca-pálida, 0.73 para pele branca, 0.75 para pele intermediária, 0.84 para pele parda e 0.98 para pele negra (BRESLIN *et al.*, 2019).

É importante ressaltar que, a despeito do desempenho satisfatório destes painéis na predição de fenótipos, por se tratar de uma metodologia ainda mais recente que a utilização de perfis *STR* - a primeira menção do termo *Forensic DNA Phenotyping* na literatura data de 2006 (KOOPS; SCHELLEKENS, 2006) -, ainda hoje nota-se resistência no cenário internacional para o uso de métodos *FDP* para auxílio em investigações. Relatórios recentes mostram que diversos países europeus não possuem legislação explícita ou implícita regulamentando a prática da *FDP*, assim como não a praticam e não apresentam discussões políticas a respeito em andamento (*i.e.:* Portugal, Dinamarca, Croácia, etc.). Por outro lado, países como Áustria, Holanda, Espanha e Suécia e Reino Unido, a prática da *FDP* já é permitida para ancestralidade biogeográfica, idade cronológica e aparência (PRAINSACK; SAMUEL, 2018). Nos Estados Unidos, apesar da ausência de regulação a nível federal e também, na maioria dos casos, estadual, a *FDP* é uma estratégia adotada por grandes companhias como *Identitas, Verogen* e *Parabon Nanolabs*. Esta última que, através da

tecnologia *Parabon Snapshot*, um sistema de fenotipagem que utiliza algoritmos de *machine lerarning* para predições de ancestralidade, fenótipos de pigmentação, morfologia facial e estimativas de parentesco a partir da análise de marcadores do tipo *SNP* (DNA FORENSICS — PARABON NANOLABS, [*s. d.*]) - projeto financiado pelo Departamento de Defesa Norte-Americano -, declara que já auxiliou na resolução de casos envolvendo a identificação de restos humanos, homicídios e de estupro. Apesar disso, não há na literatura detalhes sobre os marcadores utilizados pelo sistema, sobre o algoritmo utilizado para predição de fenótipos, ou sobre o processo de validação da metodologia, tornando-o alvo de críticas de pesquisadores influentes no campo da genética forense (ARNOLD, 2020). Entende-se, portanto, que não há um consenso internacional acerca do uso e regulação da *FDP*.

Não obstante, além do aspecto legal, ainda existem outros obstáculos a serem ultrapassados para que haja uma implementação concreta da fenotipagem forense, os quais têm sido discutidos dentro da comunidade científica e forense: a) o aspecto ético, onde debatem-se questões como o uso ou não da *FDP* como evidência em julgamentos por conta do "efeito CSI" - a confiança excessiva de um júri na tecnologia forense atual devido à ficcionalização deste tipo de evidência na cultura *pop* -, a possibilidade de violação de privacidade individual pela revelação de traços fenotípicos sensíveis (*e.g.:* predisposição à determinadas doenças genéticas), assim como o armazenamento físico destes dados genéticos e quem teria acesso à eles, temendo o uso inapropriado destas informações; b) o aspecto social, pois há uma preocupação de que a interpretação incorreta dos resultados da fenotipagem tenha um potencial para estigmatização e aumento na discriminação de grupos minoritários já considerados vulneráveis à ação do sistema criminal, visto que os traços preditos são compartilhados por grupos de pessoas; c) o aspecto genético - atualmente a principal barreira do avanço das tecnologias de fenotipagem -, considerando o conhecimento ainda muito limitado sobre as bases genéticas envolvidas na formação da morfologia facial (devido a complexidade do mapa genótipo-fenótipo), limitação que estende-se até mesmo para *EVCs* ligadas a pigmentação que contam com métodos já validados, principalmente tratando-se de fenótipos intermediários (CANALES SERRANO, 2020; SCHNEIDER; PRAINSACK; KAYSER, 2019).

1.5 IDADE CRONOLÓGICA E IDADE APARENTE

Assim como aspectos de pigmentação, a idade cronológica de um indivíduo também pode ser considerada uma *EVC* pois pode ser observada, até certo ponto, porém independente da etnicidade (HONG *et al.*, 2017). Estabelecer a idade aproximada de um indivíduo

desconhecido é de grande relevância forense e pode definir os rumos de investigações, se combinadas com informações acerca de outras *EVCs* mencionadas anteriormente.

Métodos moleculares para estimação da idade cronológica utilizando DNA de fluidos biológicos encontrados em cenas de crimes já foram propostos em literatura, estando entre os mais populares aqueles que se baseiam no perfil de metilação de sítios CpG como biomarcadores (VIDAKI; KAYSER, 2018), superando outros biomarcadores, como mRNA, comprimento telomérico e quantificação do rearranjo de DNA de células T (*sjTREC*) no que diz respeito à correlação com a idade (ZUBAKOV *et al.*, 2016). Estudos recentes mostraram uma forte associação entre marcadores de metilação em sítios CpG de diferentes genes - utilizando amostras comuns no meio forense, como sangue e saliva - e idade cronológica de indivíduos de ancestralidades distintas, apresentando modelos de predição de idade com forte correlação entre idade predita e idade cronológica, superiores a 90%, e alta acurácia de predição, com desvios médios absolutos da idade cronológica entre 3.85 e 4.25 anos (DIAS *et al.*, 2020; JUNG *et al.*, 2019).

No entanto, é preciso lembrar que, em muitos casos, há diferença entre idade cronológica de um indivíduo e o que chamamos de idade aparente ou percebida. Esta última que, diferente da idade cronológica, pode ser influenciada por fatores ambientais, como exposição prolongada à radiação ultravioleta (UV); fatores comportamentais e culturais, incluindo hábitos como tabagismo, frequência de consumo de álcool e diferentes dietas; fatores étnicos, como peles com menor pigmentação; e características faciais associadas à fatores genéticos como calvície masculina, cabelos brancos, presença de olheiras e rugas abaixo dos olhos (GUNN *et al.*, 2009; NKENGNE *et al.*, 2008; VOEGELI *et al.*, 2021).

Isto posto, é de interesse da genética forense a busca e validação de potenciais marcadores moleculares relacionados não apenas à idade cronológica de um indivíduo, mas também à sua idade aparente, por ser essa a percebida tanto ao se observar os bancos de imagens faciais disponíveis pelas equipes de investigação, quanto ao se deparar com os indivíduos durante atividades policiais externas. Contudo, alguns pontos que podem ser cobertos nos estudos eventualmente não terão o mesmo reflexo nas aplicações reais, como o uso de maquiagem, intervenções estéticas, estados hormonais transitórios (como na gravidez e durante a lactação), patologias dermatológicas, entre outros aspectos que podem modificar o estado aparente da pele de modo a alterar a percepção etária do indivíduo.

1.6 GENE *MC1R* E APARÊNCIA DE JOVIALIDADE

Em LIU et al. (2016) foram encontradas evidências da associação entre um painel de *SNPs* e *Indels* no gene *MC1R* e aparência facial em indivíduos europeus (LIU *et al.*, 2016). O estudo foi pioneiro em evidenciar que, além de fatores ambientais, variantes genéticas também podem explicar a idade aparente de um indivíduo. Para chegar a tal conclusão, o grupo de pesquisa baseou-se nos resultados de *Genome-Wide Association Studies (GWAS)* entre a idade aparente e o material genético de 2693 europeus holandeses, participantes do *Rotterdam Study*, utilizando um banco de imagens digitais de face. Foram cobertos mais de oito milhões de *SNPs* e cada imagem foi classificada por 27 avaliadores diferentes. Como resultado, observou-se que determinadas variantes próximas ou dentro do gene *MC1R*, em indivíduos de ancestralidade europeia, possuem forte associação com a idade aparente. O efeito foi mais perceptível em participantes que carregavam haplótipos de risco (R) num estado homozigoto (R/R), os quais aparentaram ser, aproximadamente, 2 anos mais velhos do que sua idade cronológica - onde R consiste na presença pelo menos um alelo de uma de quatro variantes do *MC1R* pré-selecionadas: rs1805005, rs1805007, rs1805008 e rs1805009. Tais achados foram replicados nas coortes do *Leiden Longevity Study* e *TwinsUK*, que incluem, respectivamente, dados genéticos e fotografias faciais digitais de 599 europeus holandeses e 1173 europeus. Por conta disso, as variantes selecionadas por *LIU et al.* (2016) servirão como referência para testar a hipótese deste artigo.

O gene em questão codifica o receptor de melanocortina 1 (*MC1R*), um receptor transmembrana acoplado à proteína G, presente na superfície dos melanócitos, células especializadas de pigmentação. Possuem um papel importante na regulação do processo de melanogênese nos mamíferos - determinando a quantidade e tipo de melanina produzida -, tanto em níveis basais quanto em resposta ao dano causado pela radiação UV (GARCÍA-BORRÓN; ABDEL-MALEK; JIMÉNEZ-CERVANTES, 2014). Os principais agonistas endógenos do *MC1R* são, naturalmente, as melanocortinas, um grupo de hormônios peptídicos que inclui os hormônios estimuladores de melanócitos (*MSH)* - com destaque para o ᵅ-*MSH* - e o hormônio adrenocorticoctrófico (*ACTH)*. Após a estimulação do receptor pelas melanocortinas, uma complexa cascata de sinalização é iniciada: a subunidade alfa da proteína G dissocia-se do *MC1R* e ativa a adenilato ciclase, que por sua vez catalisa a conversão de ATP em AMPc, um mensageiro secundário que irá desencadear uma série de eventos nos melanócitos (HORRELL; BOULANGER; D'ORAZIO, 2016). O aumento dos níveis de AMPc nos melanócitos estimula a síntese de eumelanina, um subtipo escuro da melanina que possui propriedades fotoprotetoras por absorção da radiação UV - diferente da

feomelanina, subtipo de coloração vermelha/amarela que confere menor proteção à radiação -, além de aumentar a eficiência de mecanismos de reparo de danos ao DNA induzidos por radiação UV (*e.g.:* dímeros de pirimidina) e diminuir o estresse oxidativo pela redução de geração de espécies reativas de oxigênio (ABDEL-MALEK *et al.*, 2009; KADEKARO *et al.*, 2005).

Certos polimorfismos com perda de função reduzem a capacidade do *MC1R* de estimular a produção de eumelanina e aumentam a proporção de feomelanina. Por conta disso, já são conhecidos por sua contribuição para fenótipos ruivos e de pele alva em um estado de heterozigose composta (VALVERDE *et al.*, 1995), caracterizados por sua tendência a queimaduras e inabilidade de bronzear-se, assim como associados ao processo de envelhecimento da pele (LAW *et al.*, 2017). Algumas destas variantes citadas na literatura estão incluídas no painel de *SNPs* e *INDELs* levantado por *LIU et al* (2016). Apesar das evidências apresentadas, mais estudos são necessários para elucidar o mecanismo celular responsável pela associação entre os *SNPs* no gene *MC1R* e um fenótipo mais jovem.

1.7 HETEROGENEIDADE ÉTNICA BRASILEIRA

Os latinoamericanos, em especial a população brasileira moderna, são caracterizados por sua heterogeneidade étnica e representam o mais recente grupo miscigenado do mundo. Por conta dessa mistura, fruto de um processo de colonização e migração (voluntária e involuntária) de diferentes povos, possuímos uma identidade fenotípica e genética bastante diversa. Estudos de estimação de ancestralidade utilizando painéis de *AIMs* indicam a presença de três grandes grupos biogeográficos, ou populações parentais, compondo a população brasileira: africanos, ameríndios e europeus (FRANCEZ; LIMA; ALMEIDA, 2015; RUIZ-LINARES *et al.*, 2014), cujo grau de influência de cada grupo entre indivíduos é variável dentro do espaço geográfico - maiores níveis de ancestralidade europeia são observados no sul do país, enquanto as ancestralidades ameríndia e africana são mais observadas no norte/oeste e nordeste/sudeste/centro-oeste, respectivamente (SALOUM DE NEVES MANTA *et al.*, 2013).

A utilização de kits comerciais de painéis de *AIMs* para estimar a ancestralidade biogeográfica (BGA) pode ser determinante para o direcionamento de investigações. Determinar aspectos étnicos de um indivíduo pode, indiretamente, fornecer informações fenotípicas úteis acerca deste. Além disso, o uso de *AIMs* também pode ser associado a painéis de *SNPs* informativos de fenótipos, tais como HIrisplex-S, já citado anteriormente. Exemplos hipotéticos da utilidade destes marcadores para o curso de investigações são

descritos por Dias Filho et al. (2020): se um determinado suspeito possui marcadores característicos que apontem uma ancestralidade africana, podemos supor que ele possui uma alta probabilidade de também ter pele escura; o mesmo vale para um possível suspeito que tenha mutações que apontem grande contribuição de ancestralidade europeia, e apresente marcadores que resultem em cabelos ruivos, em uma cidade onde só exista uma pessoa ruiva e descendente de alemães.

Vale ressaltar, no entanto, que o perfil de miscigenação da população brasileira é bastante complexo. Como mencionado anteriormente, é possível observar um menor ou maior grau de influência de determinada população parental em diferentes regiões do país, devido a variações no processo de colonização. Estudos utilizando linhagens matrilineares (DNA mitocondrial, mtDNA) e patrilineares (região não-recombinante do cromossomo Y, NRY) também mostram um processo de mestiçagem assimétrico e sexualmente enviesado, onde homens, em geral, apresentam uma maior ancestralidade europeia, enquanto uma maior contribuição africana ou ameríndia é observada em mulheres (SOUZA et al., 2019). Por conta disso, determinados estudos que buscaram caracterizar a ancestralidade brasileira concluem que, tratando-se da nossa realidade, tentar associar a ancestralidade com traços fenotípicos relevantes no âmbito forense não é uma alternativa tão aplicável ; por outro lado, Francez et al. (2015) argumenta que estes trabalhos utilizam como base *AIMs* selecionados para estudos envolvendo populações europeias, e que, quando empregados para população brasileira, superestimam o percentual de contribuição europeia enquanto minimizam a contribuição africana e ameríndia, tornando-a mais homogênea do que realmente é. Em sua tese de doutorado, Cerqueira et al. (2013) verificou que três SNPs do gene *MC1R* (rs1805007, rs1805008 e rs1805009) - presentes tanto nos painéis Irisplex, HIrisplex e HIrisplex-S, quanto no painel de SNPs associados à idade aparente elencados por Liu et al. (2016) -, não mostraram tal associação com a pigmentação nas populações gaúcha e baiana (CERQUEIRA, 2013). Estes dados sugerem que relações genótipo-fenótipo encontradas em uma população não devem ser extrapoladas para outras populações, e também a necessidade de criação de painéis específicos para populações miscigenadas.

2 JUSTIFICATIVA

Considerando a problemática exposta e a hipótese associativa de um conjunto de polimorfismos *SNP* e *Indel* no gene *MC1R* à aspectos relativos a aparência etária facial em populações europeias, compreende-se que estabelecer potenciais marcadores moleculares capazes de prever características externamente visíveis, como a classificação da percepção de idade cronológica facial (idade aparente), a partir de amostras biológicas, e que sejam aplicáveis à população miscigenada do estado do Rio grande do Sul, pode ajudar no melhor direcionamento das atividades policiais e otimizar o ritmo de investigações criminais, fornecendo subsídios à rápida e eficiente identificação de autores desconhecidos de crimes.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Reproduzir o estudo realizado por F. Liu e M. Kayser e testar a hipótese que associa um conjunto de polimorfismos *SNPs* e *Indels* no gene *MC1R* a aspectos relativos à percepção da aparência etária facial, verificando a sua aplicabilidade numa amostra representativa da população do Rio Grande do Sul.

3.2 OBJETIVOS ESPECÍFICOS

a) Extrair os dados dos *loci* polimórficos e verificar os parâmetros forenses e populacionais apresentados pelos marcadores selecionados na amostra;

b) Classificar as imagens frontais de face quanto à idade aparente através de avaliadores;

c) Estabelecer ou verificar associação entre a aparência facial percebida e polimorfismos genéticos localizados no gene *MC1R* na amostra escolhida.

4 TRABALHO EM FORMA DE ARTIGO CIENTÍFICO

(Formatado de acordo com as regras do periódico FSI: Genetics, disponíveis no Anexo A)

MC1R and age heteroclassification of face phenotypes in the Rio Grande do Sul population

Cássio Augusto Bettim[1]*, Eduardo Avila[1,2,3], Alexsandro de Vasconcellos[1,2,3], Alessandro Kahmann[3,4], Clarice Sampaio Alho[1,3]

[1]Forensic Genetics Laboratory, School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, RS, Brazil
[2]Technical Scientific Section, Federal Police Department in Rio Grande do Sul State, Porto Alegre, RS, Brazil
[3]National Institute of Science and Technology - Forensic Science, Porto Alegre, RS, Brazil
[4]Institute of Mathematics, Statistics and Physics, Federal University of Rio Grande, Rio Grande, RS, Brazil
[*]Corresponding author

cassioaugustob@hotmail.com, e.avila@edu.puc.rs, alexsandrovasconcellos@gmail.com, alessandrokahmann@furg.br, csalho@pucrs.br

1. Introduction

Forensic DNA phenotyping (FDP) consists in predicting external visible characteristics (EVCs) (*e.g.:* facial morphology, height, hair, eyes and skin pigmentation) of unknown sample donors - whether they are wanted criminals, missing or deceased persons - from crime scene DNA extracted from biological materials [1]. FDP is mainly based on Genome-Wide Association Studies (GWAS) between phenotypic traces and DNA variants, such as single nucleotide polymorphisms (SNPs) and base insertions/deletions (Indels). It can be a useful tool for providing investigative leads in cases where the Short Tandem Repeat (STR) based DNA profiling - current gold standard methodology for human identification - is not possible due to the absence of a genetic profile belonging to the suspect or no DNA matches occurrence in the available databases.

EVCs are complex traits influenced by many interacting genes, along with effects determined by environmental factors. Iris, hair, and skin pigmentation patterns are considered less complex phenotypic traits due to their high heritability, and because they are controlled by the action of fewer genes [2]. Because of this, they were the first EVCs possible to be estimated from DNA samples through prediction systems already developed and validated in the forensic community, such as Irisplex [3], HIrisplex [4] and HIrisplex-S [5] - multiplex assays targeting multiple SNPs in different genes together with statistical prediction models for eyes, hair and skin color.

Besides pigmentation traits, an individual perceived age (PA) - how old a person appears to be - can also be considered a useful EVC for human identification applied to criminal investigations. Unlike chronological age (CA), PA is influenced by numerous environmental, behavioral, cultural, ethnic, and also genetic factors: exposure to ultraviolet radiation, habits such as smoking and excessive alcohol consumption, degree of natural skin pigmentation, presence of baldness or dark circles and wrinkles below the eyes, etc. [6,7]. Molecular methods for estimating an individual CA from biological samples with high accuracy, for forensic purposes, have already been proposed in the literature [8,9,10,11], mainly using biomarkers such as CpG island methylation profile, telomere length and quantification of T-cell DNA rearrangement (sjTREC). Nevertheless, it is interesting for the forensic community to explore and validate potential molecular markers related to PA, since such phenotypic features can be observed both in the facial images databases available to the investigation teams or when encountering individuals during external police activities.

In 2016, Liu et al. [12] found evidence of the association between a panel of SNP and Indel polymorphisms in the melanocortin 1 receptor (*MC1R*) gene and PA in European subjects. The research group pioneered the identification of genetic variants that explain a small proportion of perceived facial age variation in a given population, based on GWAS studies results associated to PA, assessed from frontal and side facial images, and evaluated by 27 distinct analysts, and genomic data of 2693 dutch europeans from Rotterdam Study - a prospective cohort study ongoing since 1990 in the city of Rotterdam, in The Netherlands [12,13]. The effect was enhanced in individuals carrying risk haplotypes (R) in a homozygous state (R/R) who looked almost 2 years older than their respective CA, where R means the presence of at least one allele of four different *MC1R* variants - rs1805005, rs1805007, rs1805008 e rs1805009. These findings were replicated in the Leiden Longevity [14] Study

and TwinsUK study [15] cohorts, which include genomic data and digital face images of 599 dutch europeans and 1173 europeans, respectively.

MC1R gene encodes for a transmembrane G-protein coupled receptor (GPCR), expressed on the cell surface of melanocytes and keratinocytes, specialized pigmentation cells. It is an important element in the regulation of melanogenesis in mammals, controlling the amount and type of melanin produced both at basal levels and in response to cellular damage caused by ultraviolet (UV) radiation [44]. It has a high affinity for the group of melanocytes stimulating hormones (MSH) - especially α-MSH and adrenocorticotropic hormone (ACTH), its main endogenous agonists. Upon binding α-MSH/ACTH, MC1R activates several signalling cascades, notably cAMP pathway, inducing the synthesis of eumelanin, a dark subtype of melanin with photoprotective properties by uv radiation absorption, and reducing levels of pheomelanin, a red/yellow staining subtype that confers less radiation protection and may be involved in UV-independent pathways of oxidative stress [45,46]. Increased cAMP levels also seem to increase the efficiency of UV-induced DNA damage (e.g.: pyrimidine dimers) repair mechanisms and decrease oxidative stress by reducing the generation of reactive oxygen species [47,48].

Cerqueira et al. [16] argues, however, that care must be taken when extrapolating genotype-phenotype relationships between different populations. Some studies have shown, for example, that in addition to locus classically associated with skin pigmentation, such as SLC24A5 and SLC45A2 genes, new locus in the intergenic region of BEND7 and PRPF18 have been associated with this phenotypic trait in African admixed-populations, but not in European or Native Americans [17], while the OPRM1 and EGFR genes are associated with differences in skin pigmentation between Europeans and Native Americans [18].

As stated earlier, one of the aspects that can influence an individual PA is its ethnicity and, therefore, its biogeographic ancestry. Latin Americans, and consequently the Brazilian population, can be characterized by their ethnic heterogeneity, being considered one of the most recently admixed populations in the world [49]. It is estimated that three large parental populations make up the bulk of Brazilian genetic mosaic: Europeans, Africans and Native Americans; and the influence of each ethnic group varies within the geographical space of the country - where higher levels of European ancestry are observed in the southern region (states of Paraná, Santa Catarina and Rio Grande do Sul), while Amerindian and African ancestry are more observed in the north/west and northeast/southeast/midwest regions, respectively [19,20,21]. This multiethnic admixture is the result of an extensive process of colonization and migration (voluntary and involuntary) of different people during the country formation, which resulted in a very diverse phenotypic and genetic identity [21,22]. In this work we propose to test the hypothesis raised by Liu et al. [12] that associates a set of SNP and Indels polymorphisms of the *MC1R* gene with the difference between CA and PA in Europeans, verifying its influence on this phenotypic trait in the Rio Grande do Sul population and its applicability as a FDP marker in highly multiethnic admixed populations.

2. Materials and methods

2.1. Data collection and ethical statement

261 DNA samples from volunteer unrelated adult subjects, aged between 25 and 83 years, were selected from a human forensic biobank established in the Forensic Genetics Laboratory of Pontifical Catholic University of Rio Grande do Sul, all of them residents of Brazilian Rio Grande do Sul state. All subjects provided oral mucosa *swabs* for DNA extraction, following reading and signing consent forms (Appendix A). Participants also provided information regarding self-perceived appearance of skin color classification based on Fitzpatrick Scale, physical traits, and known familiar biogeographical origin (Appendix B). This study follows the ethical principles of Helsinki Declaration of World Medical Association [23], and was approved by the Research Ethics Committee of the Pontifical Catholic University of Rio Grande do Sul (CEP PUCRS), under number CAEE 16312319700005336.

2.2. DNA Extraction and quantification

Total DNA was manually extracted from buccal swabs using methodologies described in Avila *et al.* [24], according to phenol-chloroform-isoamyl alcohol standard protocol (Thermo Fischer Scientific Inc., MA, USA). DNA quantification was made by using Qubit™ 2.0 Fluorometer with 4 QubitTM dsDNA High Sensitivity (HS) Assay Kit (Thermo Fischer Scientific Inc., MA, USA) according to the manufacturer recommendations.

2.3. Genome sequencing

Complete genome sequencing was performed by an external service provider (*Genconve Inc., NY, USA.)* on an Illumina NextSeq 2000 (Illumina Inc., CA, USA.), using Low-Pass Whole Genome Sequencing (LP-WGS) system combined with genotype imputation step and estimated 1x full genome coverage per sample. Sequencing, data processing and library preparation on the Gencove platform were made in accordance with the company internal protocols.

2.4. Data processing and variant calling

Variant calling of the target polymorphisms in *MC1R* associated with PA in european populations - listed in the panel of Liu et al. [12] - was made using *in-house* scripts on the *VCF (variant call format)* files provided by the sequencing company for each sample, containing all sequences aligned to human genome build GRCh37, available through Gencove library.

2.5. Forensic parameters, population genetics indices and linkage disequilibrium

Forensic parameters and population genetics indices for the target polymorphisms data were computed using the browser-based aplication STRAF (version 2.0.8), being them *Random Match Probability (PM), Power of Discrimination (PD), Gene diversity (GD), Polymorphism Information Content (PIC), Power of Exclusion (PE), Typical Paternity Index (TPI), Observed Heterozigosity (Hobs),* allelic frequencies and p-value of a test for Hardy-Weinberg equilibrium (pHW) for each locus. Linkage Disequilibrium (LD) was computed using the web-based application LDlink.

2.6. Age heteroclassification

A total of 261 ICAO (International Civil Aviation Organization) standard frontal facial images of the sequenced subjects, extracted from the Brazilian national civilian identification systems, were used for age heteroclassification and determination of its PA - i.e., how old the subjects looked. Each image was classified by 11 independent volunteer evaluators (5 men and 6 women).

Prior to the assessments, a subset of ICAO standard frontal facial images from another 20 volunteer subjects without genetic data was used as a training set for the evaluators, for calibration purposes. The training set was composed of 10 men and 10 women of different age groups. Evaluators were aware of those subjects' ages. For determination of PA, we employed a previously validated methodology described in Gunn et al. [25]: the evaluators were first asked to select from a five-year range from age groups that they thought the subject looked like they belonged based on the evaluators perception, from 18 to 89 years; secondly, the evaluators were asked for a precise age within the selected age group.

After the training session, evaluators scored the PA for the 261 sequenced subjects, for whom they were unaware of the subjects chronological age or any other personal information. Data for each evaluator were recorded in a spreadsheet previously prepared by the authors of this study.

2.7. Statistical analysis

As for the statistical analysis, Multiple Linear Regression (MLR) was employed to model the influence of polymorphisms (SNPs and Indels) in the difference in years between the CA and the PA (Δage). The MLR equation estimated by the aforementioned data provides a numerical description of the influence of independent variables (polymorphisms) in the variation of the response variable (Δage) [26]. Since the independent variables are categorical, the regression equation estimates the mean difference between a reference level and another level of the variable. In this paper we stated as reference level the most common level, i.e., the level with more observations. Additionally, MLR assumes that all independent variables are independent among themselves (i.e., the inexistence of multicollinearity), which usually is not the case for genetic data [26,27]. Therefore, in order to provide a parsimonious model, and reduce the multicollinearity issue, two approaches were employed: i) SNPs with less then 4 observations in the non-reference levels were excluded, since such SNPs would not present enough variability to actually add information to the model; and ii) a stepwise variable selection based on the Akaike Information Criteria (AIC) was performed in the model, aiming to avoid multicollinearity. All statistical analyses were performed in *R* (version 4.0.5) and the packages “stats” and “MASS”, for MLR and variable selection respectively.

3. Results and discussion

Variant calling from sequencing data resulted in the identification of 12 SNPs and 52 Indels, totaling 64 polymorphisms. Five polymorphisms were selectively removed because: i) all 261 individuals in the sample had the same genotype in the given locus, making it impossible to verify its influence in the PA; ii) were not identified by automatic variant calling. More detailed data about the 59 selected polymorphisms can be seen in Table S1. Interestingly, after verification in the dbSNP and different from that presented in the study by Liu et al [12], we noticed that variants associated with apparent facial age were not limited to the *MC1R* gene region, but also spread in adjacent genes in 16q24.3 region, such as *AFG3L1P*, *DEF8*, *FANCA*, *TUBB3*, *SPATA2L*, *SPIRE2*, *VPS9D1*, and *ZNF276*. Regarding the age heteroclassification of the sample from the frontal facial images, we obtained a total of 2871 PA assessments. Characteristics of the total sample, and also divided among the different skin types constituents, are presented in Tables 1 and 2 respectively. When analyzed individually, it was not possible to observe the influence of the variants and the presence of effect alleles (R) on the Δage (figure 1). This corroborated the need for a multivariate analysis of the data.

In our sample of 261 inhabitants of Rio Grande do Sul, MLR found statistically significant relationships between 9 SNPs and Indel variants in different genes and PA. Specifically, rs12924124 (*ZNF276*), rs12931267 (*FANCA*), rs36100920 (*SPIRE2*/*LOC105371419*), rs59751572 (*TUBB3*), rs201156703 (*AFG3L1P*), rs112220510 (*AFG3L1P*), rs112373300 (*AFG3L1P*), rs112388869 (*TUBB3*), and rs34850152 (*AFG3L1P*). The results of the MLR can be seen in more detail in Table 3.

Interestingly, in the Rio Grande do Sul population sample, and differently from that observed in European populations, the results of the MLR shows that many R alleles from different variants in *AFG3L1P*, *TUBB3*, and *FANCA* genes influenced PA so that individuals appeared to be 1,5-12 years younger, approximately, and being significantly associated ($p<0.05$) with a younger face appearance (Table 3). For example, *TUBB3* rs59751572 and *AFG3L1P* rs112220510 showed an decrease in PA per increase in R - β values (variation in Δage in relation to the reference genotype) of -1,26±0,54 for heterozygotes and -3,16±1,34 for homozygotes; -1,48±0,67 for heterozygotes and -6,09±1,83 for homozygotes, respectively. The opposite effect was observed for carriers of variants *ZNF276* rs12924124, *SPIRE2*/*LOC105371419* rs36100920, *AFG3L1P* rs34850152, and *AFG3L1P* rs201156703, where the presence of R allele made the subjects appear to be 6-12 years older than non carriers, approximately (Table 3), being significantly associated ($p<0.05$) with individuals with an older than expected facial appearance.

We also performed MLR analysis after dividing the sample into two groups: the light-skinned individuals (types I, II, and III) and the dark-skinned individuals (IV, V, and VI), in order to identify the influence of the variants extracted in PA within the two groups presenting distinct ethnicity background. For the light-skinned group, the presence of R alleles of *AFG3L1P* rs112220510, *AFG3L1P/CENPBD1* rs151289005, *TUBB3* rs59751572 and *DEF8* rs113500176 were significantly associated ($p<0.05$) with individuals with an 2-7 years younger face appearance (Table 4), while R alleles of *AFG3L1P* rs113020559, *AFG3L1P* rs34850152 and *SPIRE2/LOC105371419* rs36100920, were significantly associated ($p<0.05$) with individuals with an 2-13 years older face appearance. In the dark-skinned

group, the presence of R alleles in the selected variants *AFG3L1P* rs112556696, *AFG3L1P* rs112220510, *TUBB3* rs59751572, *TUBB3* rs150764400, *DEF8* rs113500176, *DEF8* rs1112881095, and *DEF8* rs11453006 showed to be significantly associated ($p<0.05$) with a 4-18 years younger appearance (Table 5), while *AFG3L1P* rs113020559, *TUBB3* rs372052072, *DEF8* rs111421479, *DEF8* rs11446223, were associated with an older facial appearance, ranging from 3-12 years, approximately.

Forensic parameters obtained for the analyzed genetic variants are presented in Table S2. Resulting values suggest that this set of markers does not possess sufficient genetic information to be successfully used in human identification (when considered the individualization aspect only). Such a condition is caused mainly by the low diversity found in SNP or Indel variants (where biallelic forms are the most common in this kind of locus), which can be a major factor for the reduced heterozygosity presented by such markers. In addition, minor alleles in most markers also present a significant difference in frequency in comparison to the major alleles, once again reducing the utility of these variants in human genetic identification. High heterozygosity levels are desired for human identification purposes, and biallelic markers to be used in such applications should be carefully selected to maximize the panel efficiency [23]. P-values of Hardy-Weinberg Equilibrium (HWE) test, referred as pHW (Table S2), indicate that certain variants - some of them among those selected by MLR, such as rs12924124, rs12931267, rs201156703, rs36100920 - present a significant deviation ($p<0.05$) from HWE. A more likely cause for this finding is the existence of some degree of genetic stratification within the target population, which are being detected with deviations from HWE. Additional testing must be performed to better characterize the occurrence of genetic substructures associated with the evaluated sample, but the presence of people from distinct biogeographical origins is probably linked to this phenomenon. Linkage disequilibrium (LD) analysis also showed a high level of correlation (R^2) between these same variants (Figures 1,2, and 3), indicating non-independent segregation of the loci during the formation of haplotypes.

None of the four polymorphisms in the *MC1R* gene highlighted by Liu et al [12] for their relationship with PA in Europeans - rs1805005, rs1805007, rs1805008 and rs1805009 - showed to have such association in our study. These SNPs are classically associated with red hair, fair skin and freckles phenotypes - which is why they are also called RHC (Red Hair Color) variants - and also reduced photoprotection in European populations [29,30,43], but they did not show to have the same influence in Latin Americans, besides being quite rare in these populations [31,16]. In fact, in addition to the few redheads individuals in our sample, these SNPs presented low allelic frequencies (Table S1), which should explain that they were not selected by the MLR model.

β values found in this study are considerably higher than the effect on PA observed in the study by Liu et al [12]. A possible explanation for this may be that our study used a smaller sample and number of evaluators, and consequently, a smaller - but still significant - number of total observations, in addition to presenting a greater amplitude of age groups, composed mostly of relatively young individuals (more than 50% had a CA lower than 35 years), and therefore, a considerably lower mean age (Table 1). The estimated age of an individual is a very subjective aspect that can be influenced not only by the facial characteristics of the target person - that depend on environmental, cultural and behavioral,

ethnic and genetic factors (*e.g.*: lifelong sun exposure, different diets, smoking, presence of dark circles, male baldness etc.) [6,7,32,33] - , but also by factors such as the difference in gender and age group between evaluator and research subject and facial expression of the target person [7,32,34]. In addition, there is no standard methodology for PA determination. It is also important to point out that because certain variants are in LD and therefore correlated with each other, it can be interpreted that their effects can be summed up or may be canceling each other in the regression model (*e.g.:* rs12924124 and rs12931267 are in LD and presented β values of +10,60 and -11,86, respectively, which means a resulting β value of -1,26 years in PA in individuals with both variants). This additive model can also explain why some genetic variants seem to have, according to the model, a very drastic effect in estimated PA, with age variations over 10 years (older or younger). Despite the fact that the influence of a single variant may be overestimating its role on PA determination, when considered together with other genetic markers its effect is balanced. Therefore, it is important to highlight that the presented variantes only make sense when considered together in the context of the statistical model, and should not be individually considered alone in any attempt to pinpoint genetic polymorphisms associated with facial appearance.

Regarding the age heteroclassification, individuals with higher skin pigmentation (skin types V and VI) showed a lower mean PA (34,9 years) than the mean CA (36 years), indicating that this group, in general, appears to have a younger face phenotype, something that is not observed in the other skin types. These results are in consensus with the literature, since individuals with lighter skin have greater sensitivity to UV-induced damage because they have lower levels of eumelanin in the epidermis [28] and being more conducive to the appearance of signs of aging. Our results also show a clear difference in variants significantly associated with face phenotypes selected by the regression model for the light-skinned and dark-skinned groups (see Tables 4 and 5). We believe that this phenomenon may be caused by the significant difference in the presence of individuals with darker skin (types IV, V and VI) and lighter skin (types I, II and III) in the sample, since: i) the dark-skinned group represents few less than 22% of the total sample (Table 1); ii) the vast majority of the selected variants do not present a higher frequency in individuals of African ancestry; iii) it is possible to observe the presence of selected variants in common in the panels relative to the total sample (Table 3) and the light-skinned group (Table 5), which represents the largest part of the sample (e.g.: rs36100920, rs59751572 and rs34850152).

Unfortunately, the molecular mechanisms behind the effect of these variants on PA remain unknown. Of all the selected variants that showed a significant association with younger or older face phenotypes in our study - with the exception of FANCA rs12931267 -, including those in the light-skinned and dark-skinned groups, none presents clinical significance or association with any other phenotypic trait reported in ClinVar [36] and GWAS catalog [37] databases, nor was it directly cited in previous publications. *FANCA* rs12931267 has been associated with red hair phenotypes and freckling in European and Canadian populations [35,36]. Most of the selected markers are intronic variants (Tables 3,4 and 5), with the exception of: i) *LOC105371419* rs36100920, 3'-UTR (untranslated region) variant; ii) *AFG3L1P* rs112220510, downstream transcript variant, iii) *AFG3L1P/CENPBD1* rs151289005, upstream/stop gained variant; and iv) *SNORA119* rs111421479, upstream variant.

A significant portion (43%) of trait-disease associated loci SNPs are found in non-coding regions of the genome such as introns, responsible for regulating gene expression, and can induce events such as exon skipping, activation of cryptic splicing sites and altering the balance of isoforms that are an alternative splicing products [37,38,39,40]. In fact, the 16q24.3 region has a high gene density, which adds to an inefficient polyadenylation tail in the *MC1R* gene, allows the intergenic splicing of this locus with adjacent genes, for example TUBB3 [41,42]. This intergenic splicing results in an *MC1R-TUBB3* chimera, which is transcribed in an non-canonical mRNA that is translated in two functional isoforms of canonical *MC1R* - Iso1 and Iso2 [41,42]. These isoforms preserve the general structure of an GPCR, being able to perform *MC1R* functions, and their expression appears to be induced by the action of α-MSH, the *MC1R* agonist [42,43]. Despite having a poor ability to activate cAMP pathway like the canonical *MC1R* due to its decreased expression on the cell surface, it was proposed that Iso1 and Iso2 may be involved in fine-tune of pigmentation response by playing a role in the rearrangement of the cytoskeleton that is required to initiate, propagate and guide the dendrites of activated melanocytes [41,42]. Considering these points, a possible hypothesis we can raise is that the presence of the selected intronic variants in genes close to the *MC1R* locus may be affecting potential regulatory elements and intergenic splicing sites. This context could lead to a decrease in gene expression of both the canonic *MC1R* and potential functional isoforms, such as Iso1 and Iso2, and consequently in the photoprotective mechanisms involving its signaling pathways. Even if this theory could explain why certain variants influence an older face phenotype in comparison to the reference allele (β positive values), the same cannot be said for cases where individuals appeared to be younger than they actually are. Further studies will be needed so that the molecular mechanisms behind the influence of these SNPs and Indel polymorphisms on PA can be better elucidated.

4. Conclusions

In summary, our study was successful in demonstrating the influence of certain SNPs and Indel variants - included in the panel raised by Liu et al. [12] -, present in different genes adjacent to the *MC1R* gene in 16q24.3 region, to the PA in the population of Rio Grande do Sul. Interestingly, it was observed that, unlike what was evidenced in European populations, such variants seem to contribute to both younger and older face phenotypes in our population. Also, since few polymorphisms among all the extracted variants showed such an effect in our sample, our results suggest the need for additional studies of genotype-phenotype association with regard to apparent age in the Rio Grande do Sul population, seeking to find new candidate genes that may be associated with this complex trait and targeting the developing of forensic marker panels specifically for highly mixed populations. Despite our findings, further studies are still needed to understand the molecular basis behind the influence of these polymorphisms on complex traits like face phenotypes of different population groups.

5. Acknowledgments

Authors would like to thank the Human Forensic Genetic Laboratory staff who assisted in all phases in this work. and the financial support received by Brazilian funding agencies CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Pesquisa Científica), and FAPERGS (Fundação de Amparo à Pesquisa do Rio Grande do Sul).

References

[1] Kayser M. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet. 2015 Sep;18:33-48. doi: 10.1016/j.fsigen.2015.02.003. Epub 2015 Feb 16. PMID: 25716572.

[2] 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(3):154-161. doi:10.1016/j.fsigen.2009.01.012

[3] 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. Forensic Sci Int Genet. 2011 Nov;5(5):464-71. doi: 10.1016/j.fsigen.2010.09.008. Epub 2010 Oct 14. PMID: 20947461

[4] Walsh S, Liu F, Wollstein A, Kovatsi L, et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet. 2013. 7: 98-115. ISSN 1872-4973. doi: 10.1016/j.fsigen.2012.07.005

[5] Chaitanya L, Breslin K, Zuñiga S, et al. The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci Int Genet. 2018;35:123-135. doi:10.1016/j.fsigen.2018.04.004

[6] Gunn DA, Rexbye H, Griffiths CE, et al. Why some women look young for their age. PLoS One. 2009;4(12):e8021. Published 2009 Dec 1. doi:10.1371/journal.pone.0008021

[7] Nkengne A, Bertin C, Stamatas GN, et al. Influence of facial skin attributes on the perceived age of Caucasian women. J Eur Acad Dermatol Venereol. 2008;22(8):982-991. doi:10.1111/j.1468-3083.2008.02698.x

[8] Vidaki A, Kayser M. Recent progress, methods and perspectives in forensic epigenetics. Forensic Sci Int Genet. 2018;37:180-195. doi:10.1016/j.fsigen.2018.08.008

[9] Zubakov D, Liu F, Kokmeijer I, et al. Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. Forensic Sci Int Genet. 2016;24:33-43. doi:10.1016/j.fsigen.2016.05.014.

[10] Dias HC, Cordeiro C, Pereira J, et al. DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay. Forensic Sci Int. 2020;311:110267. doi:10.1016/j.forsciint.2020.110267

[11] Jung SE, Lim SM, Hong SR, Lee EH, Shin KJ, Lee HY. DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. Forensic Sci Int Genet. 2019;38:1-8. doi:10.1016/j.fsigen.2018.09.010

[12] Liu F, Hamer MA, Deelen J, et al. The MC1R Gene and Youthful Looks. Curr Biol. 2016;26(9):1213-1220. doi:10.1016/j.cub.2016.03.008

[13] Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol. 2017;32(9):807-850. doi:10.1007/s10654-017-0321-4

[14] Schoenmaker M, de Craen AJ, de Meijer PH, et al. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet. 2006;14(1):79-84. doi:10.1038/sj.ejhg.5201508

[15] Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK Adult Twin Registry (TwinsUK Resource). Twin Research and Human Genetics. 2013;16(1):144-149. doi:10.1017/thg.2012.89

[16] Cerqueira CCS. Genotype-phenotype relationship and human pigmentation: evolutionary aspects and application in forensic genetics. Doctoral thesis presented in the Postgraduate Program in Genetics and Molecular Biology - Federal University of Rio Grande do Sul, Porto Alegre, RS. 2013.

[17] Hernandez-Pacheco N, Flores C, Alonso S, et al. Identification of a novel locus associated with skin colour in African-admixed populations. Sci Rep. 2017;7:44548. Published 2017 Mar 16. doi:10.1038/srep44548

[18] Quillen EE, Bauchet M, Bigham AW, et al. OPRM1 and EGFR contribute to skin pigmentation differences between Indigenous Americans and Europeans. Hum Genet. 2012;131(7):1073-1080. doi:10.1007/s00439-011-1135-1

[19] Francez PA, Ribeiro-Rodrigues EM, dos Santos SE. Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. Forensic Sci Int Genet. 2012;6(1):132-135. doi:10.1016/j.fsigen.2011.04.002

[20] Saloum de Neves Manta F, Pereira R, Vianna R, et al. Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. PLoS One. 2013;8(9):e75145. Published 2013 Sep 20. doi:10.1371/journal.pone.0075145

[21] Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, et al. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genet. 2014;10(9):e1004572. Published 2014 Sep 25. doi:10.1371/journal.pgen.1004572

[22] Pena SD, Bastos-Rodrigues L, Pimenta JR, Bydlowski SP. DNA tests probe the genomic ancestry of Brazilians. Braz J Med Biol Res. 2009;42(10):870-876. doi:10.1590/s0100-879x2009005000026

[23] World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA. 2013;310(20):2191–2194. doi:10.1001/jama.2013.281053

[24] Avila E, Speransa PA, Lindholz CG, Alho, CS. Haplotype distribuition in a forensic full mtDNA genome database of admixed Southern Brazilians and its association with self-declared ancestry and pigmentation traits. Forensic Sci Int Genet. 2021. 57: 102650. doi: 10.1016/j.fsigen.2021.102650

[25] Gunn DA, Murray PG, Tomlin CC, Rexbye H, Christensen K, Mayes AE. Perceived age as a biomarker of ageing: a clinical methodology. Biogerontology. 2008;9(5):357-364. doi:10.1007/s10522-008-9141-y

[26] Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. Hoboken, NJ: Wiley; 2021.

[27] Han YU, Jiang S, Land KC. Multicollinearity in hierarchical linear models. Social sciences research. 2015 Sep. 53:118-136. doi: 10.1016/j.ssresearch.2015.04.008

[28] D'Orazio J, Jarrett S, Amaro-Ortiz A, Scott T. UV radiation and the skin. Int J Mol Sci. 2013;14(6):12222-12248. Published 2013 Jun 7. doi:10.3390/ijms140612222

[29] Haddadeen C, Lai C, Cho SY, Healy E. Variants of the melanocortin-1 receptor: do they matter clinically?. Exp Dermatol. 2015;24(1):5-9. doi:10.1111/exd.12540

[30] Makova K, Norton H. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. Peptides. 2005;26(10):1901-1908. doi:10.1016/j.peptides.2004.12.032

[31] 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 in humans. Nat Genet. 1995;11(3):328-330. doi:10.1038/ng1195-328

[32] Bulpitt CJ, Markowe HL, Shipley MJ. Why do some people look older than they should?. Postgrad Med J. 2001;77(911):578-581. doi:10.1136/pmj.77.911.578

[33] Clatici VG, Racoceanu D, Dalle C, et al. Perceived Age and Life Style. The Specific Contributions of Seven Factors Involved in Health and Beauty. Maedica (Bucur). 2017;12(3):191-201.

[34] Voelkle MC, Ebner NC, Lindenberger U, Riediger M. Let me guess how old you are: effects of age, gender, and facial expression on perceptions of age. Psychol Aging. 2012;27(2):265-277. doi:10.1037/a0025065

[35] Lona-Durazo F, Mendes M, Thakur R, et al. A large Canadian cohort provides insights into the genetic architecture of human hair colour. Commun Biol. 2021;4(1):1253. Published 2021 Nov 4. doi:10.1038/s42003-021-02764-0

[36] Eriksson N, Macpherson JM, Tung JY, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet. 2010;6(6):e1000993. Published 2010 Jun 24. doi:10.1371/journal.pgen.1000993

[37] Lin H,, Hargreaves KA, Li R, et al. RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants. Genome Biol 20, 254 (2019). doi:10.1186/s13059-019-1847-4

[38] Pagani F, Baralle F, Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5, 389–396 (2004). doi:10.1038/nrg1327

[39] Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet. 2016;17(1):19-32. doi:10.1038/nrg.2015.3

[40] Douglas AG, Wood MJ. RNA splicing: disease and therapy. Brief Funct Genomics. 2011;10(3):151-164. doi:10.1093/bfgp/elr020

[41] Dalziel M, Kolesnichenko M, das Neves RP, Iborra F, Goding C, Furger A. Alpha-MSH regulates intergenic splicing of MC1R and TUBB3 in human melanocytes. Nucleic Acids Res. 2011;39(6):2378-2392. doi:10.1093/nar/gkq1125

[42] Herraiz C, Olivares C, Castejón-Griñán M, Abrisqueta M, Jiménez-Cervantes C, García-Borrón JC. Functional Characterization of MC1R-TUBB3 Intergenic Splice Variants of the Human Melanocortin 1 Receptor. PLoS One. 2015;10(12):e0144757. Published 2015 Dec 11. doi:10.1371/journal.pone.0144757

[43] Manganelli M, Guida S, Ferretta A, et al. Behind the Scene: Exploiting MC1R in Skin Cancer Risk and Prevention. Genes (Basel). 2021 Jul;12(7): 1093. doi:10.3390%2Fgenes12071093

[44] García-Borrón JC, Abdel-Malek Z, Jiménez-Cervantes C. MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. Pigment Cell Melanoma Res. 2014;27(5):699-720. doi:10.1111/pcmr.12257

[45] Panzella L, Leone L, Greco G, et al. Red human hair pheomelanin is a potent pro-oxidant mediating UV-independent contributory mechanisms of melanomagenesis. Pigment Cell Melanoma Res. 2014;27(2):244-252. doi:10.1111/pcmr.12199

[46] Tanaka H, Yamashita Y, Umezawa K, Hirobe T, Ito S, Wakamatsu K. The Pro-Oxidant Activity of Pheomelanin is Significantly Enhanced by UVA Irradiation: Benzothiazole Moieties Are More Reactive than Benzothiazine Moieties. Int J Mol Sci. 2018;19(10):2889. Published 2018 Sep 23. doi:10.3390/ijms19102889

[47] Swope V, Alexander C, Starner R, Schwemberger S, Babcock G, Abdel-Malek ZA. Significance of the melanocortin 1 receptor in the DNA damage response of human melanocytes to ultraviolet radiation. Pigment Cell Melanoma Res. 2014;27(4):601-610. doi:10.1111/pcmr.12252

[48] Kadekaro AL, Kavanagh R, Kanto H, et al. alpha-Melanocortin and endothelin-1 activate antiapoptotic pathways and reduce DNA damage in human melanocytes. Cancer Res. 2005;65(10):4292-4299. doi:10.1158/0008-5472.CAN-04-4535

[49] Dias Filho CR, Rodrigues EL, Francez PAC. Introduction to forensic genetics. Campinas, SP: Millenium; 2020.

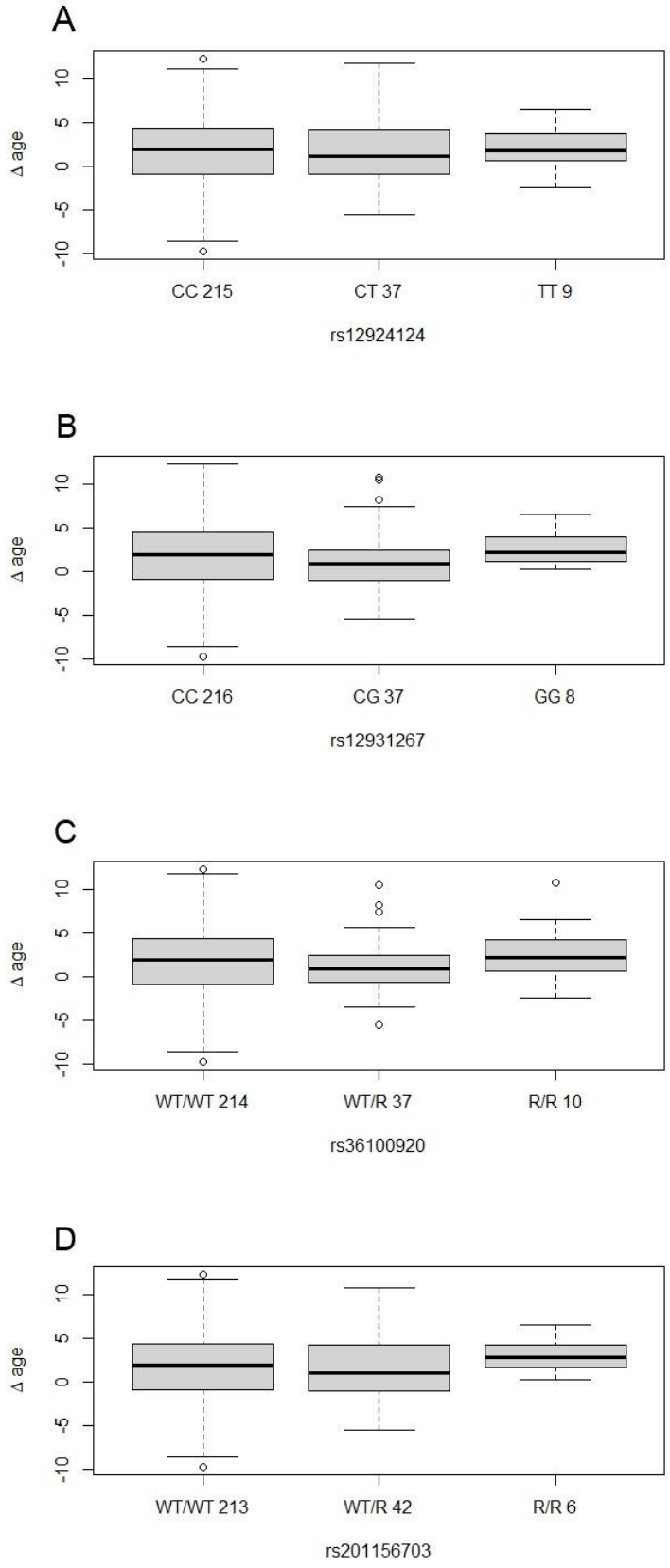


Figure 1:Boxplot of Δage in WT/WT, WT/R and R/R genotypes in four different variants. A) rs12931267. **B)** rs12931267. **C)** rs36100920. **D)** rs201156703. WT - Wild type allele, R - Effect allele. Data represented as median and percentile.

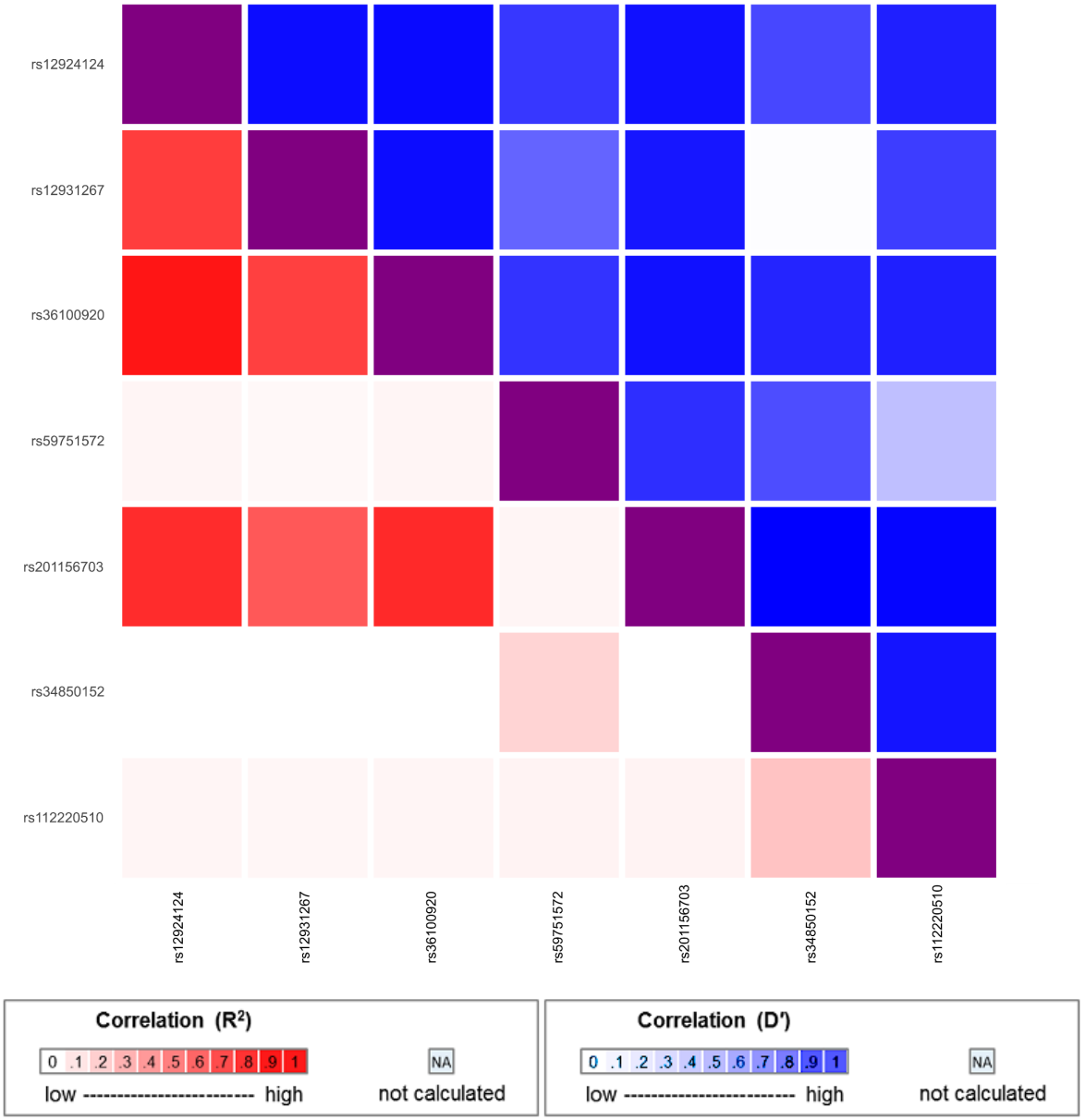


Figure 2:Linkage Disequilibrium (LD) matrix of the variants selected by MLR. Variants rs112373300 rs112388869 are missing from 1000G (GRCh37) data.

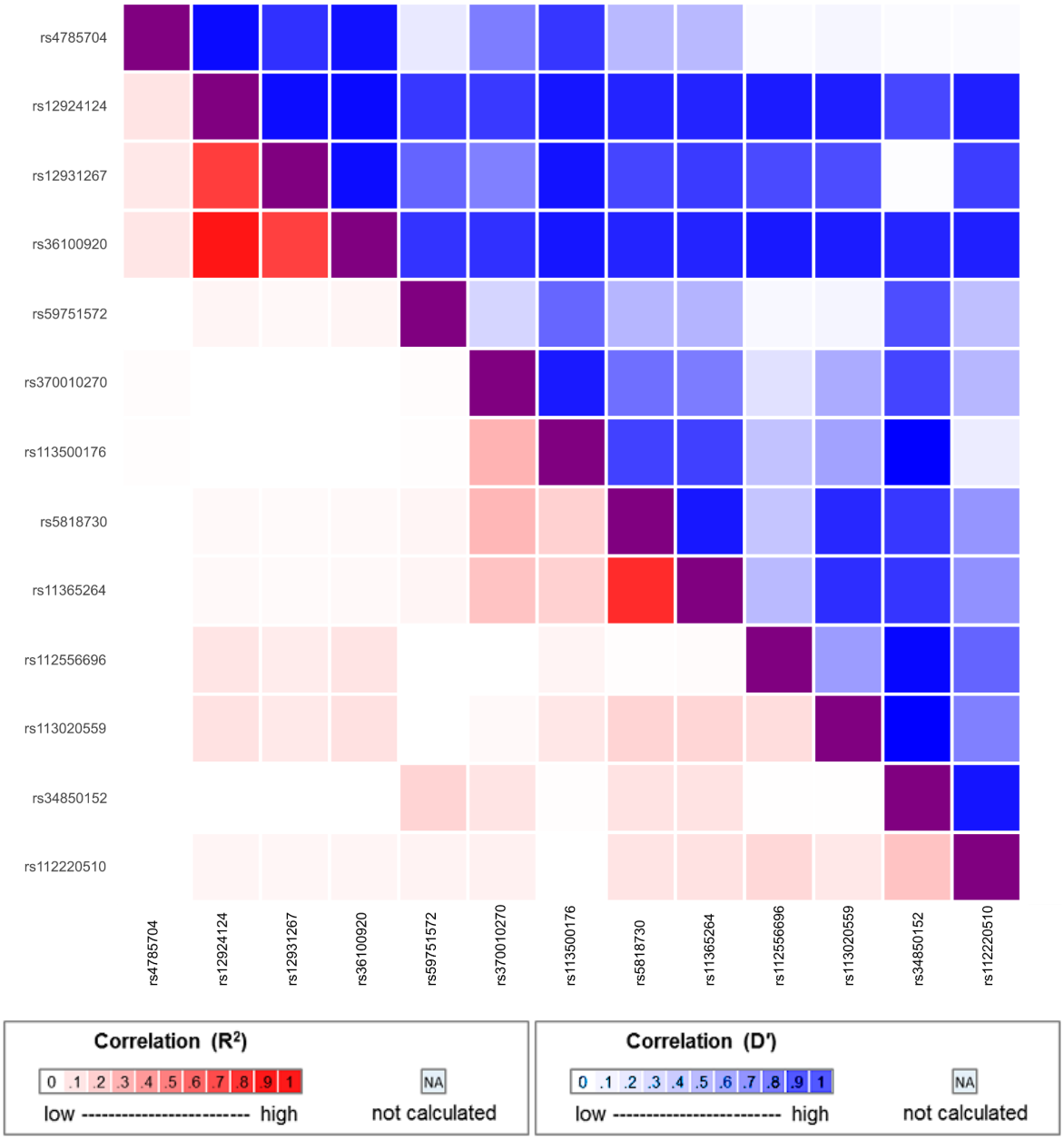


Figure 3: Linkage Disequilibrium (LD) matrix of the variants selected by MLR in light-skinned group. Variants rs112388869, rs11446223 and rs151289005 are missing from 1000G (GRCh37) data.

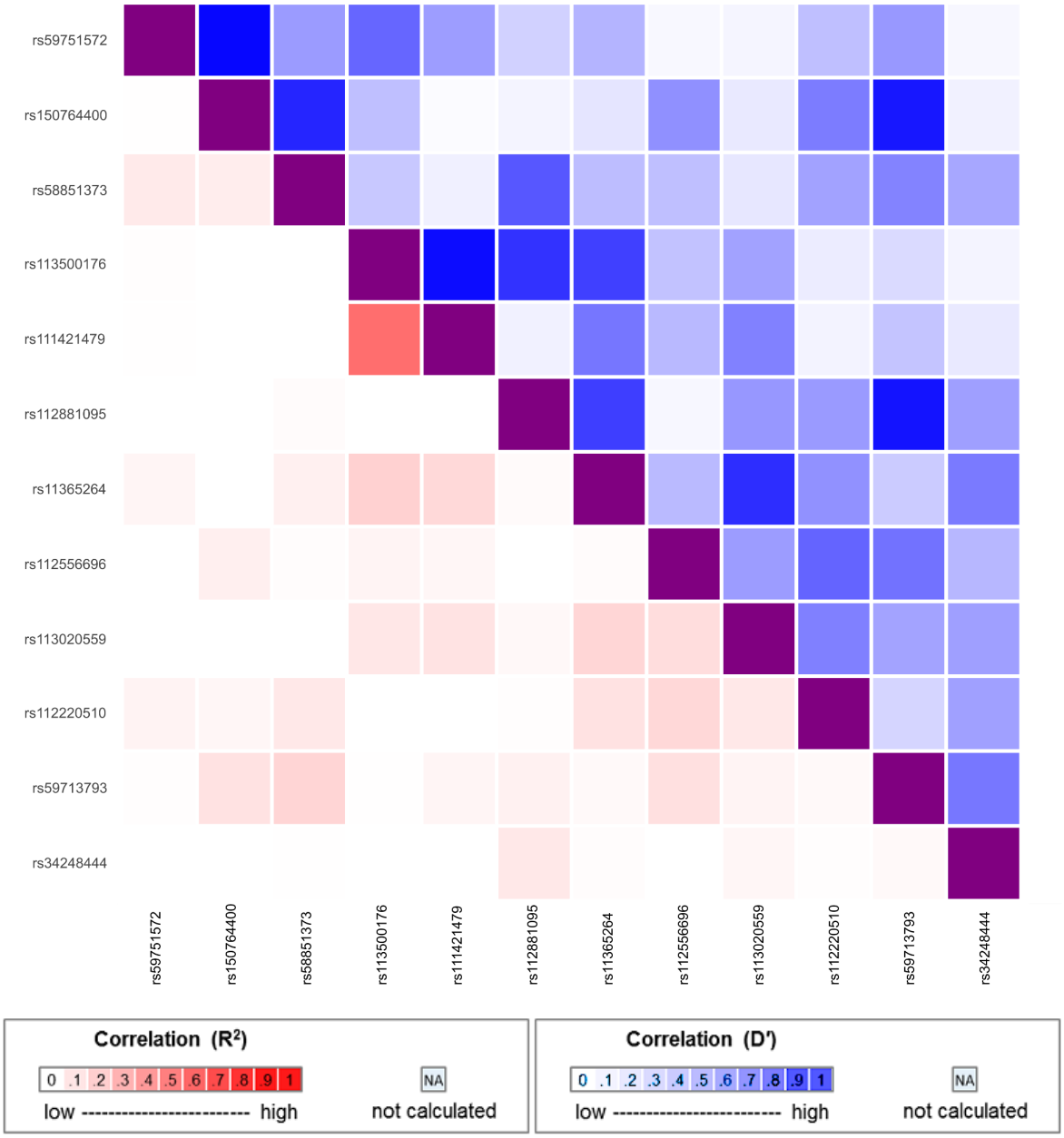


Figure 4: Linkage Disequilibrium (LD) matrix of the variants selected by MLR in dark-skinned group. Variants rs112388869, rs11446223 and rs11453006 are missing from 1000G (GRCh37) data.

Table 1. Characteristics of the sample

Characteristics	**Sample**	
Male (N,%)	111	42.53
Female (N,%)	150	57.47
CA (Mean,SD)	33	10.29
PA (Mean,SD)	35.54	10
ΔAge(Mean±SD)	1.72	4
Skin type I (N)	32	-
Skin type II (N)	137	-
Skin type III (N)	35	-
Skin type IV (N)	16	-
Skin type V (N)	25	-
Skin type VI (N)	16	-

N - Number of individuals; SD - Standard deviation; CA - Chronological Age; PA - Perceived Age; ΔAge - difference between PA and CA

Table 2. Characteristics of the sample separated by skin types.

Characteristcs	**Skin type I-II**	**Skin type III-IV**	**Skin type V-VI**
CA (Mean±SD)	33±10.13	34±11.45	36±9.49
PA (Mean±SD)	34.9±9.79	36.81±10.9	34.9±9.7
ΔAge (Mean±SD)	1.63±4.03	1.81±4,21	2.18±3.67

SD - Standard deviation; CA - Chronological Age; PA - Perceived Age; ΔAge - difference between PA and CA

Table 3. SNP and Indel variants that most influenced in PA.

Locus	Gene: Consequence	Genotype	β	SD	p-value
rs12924124	ZNF276: Intron variant	TT	2.04	6.75	0.76204
		CT	10.60	3.79	**0.005575**
rs12931267	FANCA: Intron Variant	CG	-11.13	4.15	**0.007869**
		GG	-11.86	6.32	0.061642
rs36100920	SPIRE2: Intron variant, LOC105371419: 3'-UTR variant	WT/R	0.50	1.59	0.753815
		R/R	11.87	4.29	**0.006142**
rs59751572	TUBB3: Intron variant	WT/R	-1.26	0.59	**0.0323**
		R/R	-3.16	1.34	**0.019211**
rs201156703	AFG3L1P: Intron variant	WT/R	1.64	1.22	0.180801
		R/R	7.54	3.59	**0.036815**
rs112220510	AFG3L1P: Intron variant, downstream transcript variant	WT/R	-1.48	0.67	**0.028644**
		R/R	-6.09	1.83	**9.92E-04**
rs112373300	AFG3L1P: Intron variant	WT/R	0.93	0.64	0.150204
		WT/WT	-5.55	2.68	**0.03913**
rs112388869	TUBB3: Intron variant	WT/R	-2.52	0.99	**0.010972**
rs34850152	AFG3L1P: Intron variant	WT/R	4.93	1.32	**2.26E-04**

WT - Wild type allele; R - Effect allele; β - variation in Δage in relation to the reference genotype (most common genotype); SD - standard deviation.

Table 4. Variants that most influenced in PA on individuals of light-skinned group.

Locus	Gene: Consequence	Genotype	β	SD	p-value
rs4785704	NA	AG	2.27	1.41	**1.09E-01**
		GG	5.78	8.33	0.48855
rs12924124	ZNF276: Intron variant	CT	7.17	4.20	0.0899
		TT	NA	NA	NA
rs12931267	FANCA: Intron Variant	CG	-8.21	4.38	0.06279
		GG	-15.55	8.09	0.05632
rs112556696	AFG3L1P: Intron variant	AG	2.22	1.51	0.14342
		GG	-6.04	5.79	0.29893
rs36100920	SPIRE2: Intron variant, LOC105371419: 3'-UTR variant	WT/R	0.05	1.97	0.97874
		R/R	9.61	4.55	**0.03602**
rs59751572	TUBB3: Intron variant	WT/R	-1.58	0.69	**0.02235**
		R/R	-1.48	1.53	0.33537
rs370010270	DEF8: Intron variant	WT/R	1.65	0.78	**0.03677**
		R/R	2.27	1.65	0.17049
rs113500176	DEF8: Intron variant	WT/R	-0.83	0.93	0.37035
		R/R	-4.84	2.17	**0.02713**
rs5818730	AFG3L1P: Intron variant	R/R	3.53	2.32	0.13
		WT/WT	-3.67	2.74	0.18234
rs11365264	AFG3L1P: Intron variant	R/R	-4.25	2.27	0.06242
		WT/WT	3.68	2.69	0.17287

Locus	Gene: Consequence	Genotype	β	SD	p-value
rs113020559	AFG3L1P: Intron variant	WT/R	1.14	1.90	0.54786
		R/R	12.70	6.10	**0.03873**
rs72468543	NA	WT/R	-2.45	1.32	0.06456
		R/R	2.85	3.68	0.43948
rs112220510	AFG3L1P: Intron variant, downstream transcript variant	WT/R	0.02	1.08	0.98441
		R/R	-6.53	2.47	**9.01E-03**
rs112388869	TUBB3: Intron variant	WT/R	-2.64	1.35	0.05253
rs11446223	DEF8: Intron variant	WT/R	-1.79	1.38	0.19608
		R/R	3.50	4.18	0.40316
rs151289005	CENPBD1: Stop gained, AFG3L1P: Upstream variant	WT/R	-5.52	2.53	**3.02E-02**
rs34850152	AFG3L1P: Intron variant	WT/R	3.77	1.66	**0.02422**

WT - Wild type allele; R - Effect allele; β - variation in Δage in relation to the reference genotype (most common genotype); SD - standard deviation; NA - Non available.

Table 5. Variants that most influenced in PA on individuals of dark-skinned group.

Variant	Gene: Consequence	Genotype	β	SD	p-value
rs112556696	AFG3L1P: Intron variant	AG	-3.65	1.43	**0.017361**
		GG	-17.39	3.82	**0.000108**
rs372052072	TUBB3: Intron variant	WT/R	3.06	1.41	**0.039396**
		R/R	-3.32	2.24	0.150596
rs59751572	TUBB3: Intron variant	WT/R	-6.01	1.69	**0.001433**
rs150764400	TUBB3: Intron variant	WT/R	-6.25	1.50	**0.000296**
rs58851373	NA	WT/R	5.50	1.76	**0.004442**
		R/R	9.56	2.44	**0.000593**
rs113500176	DEF8: Intron variant	WT/R	-7.37	1.63	**0.000118**
rs111421479	DEF8: Intron Variant, SNORA119: Upstream variant	WT/R	6.16	1.64	**0.000871**
		R/R	-3.01	3.55	0.404337
rs112881095	DEF8: Intron variant	WT/R	-5.19	1.87	**0.010178**
rs11365264	AFG3L1P: Intron variant	R/R	-0.90	1.28	0.489766
		WT/WT	-2.12	1.55	0.183195
rs10611547	NA	WT/R	-0.97	1.18	0.415563
		WT/WT	5.18	4.76	0.286385
rs113020559	AFG3L1P: Intron variant	WT/R	-0.42	1.28	0.745114
		R/R	8.11	3.65	**0.035366**
rs72468543	NA	WT/R	7.53	1.71	**1.65E-04**
rs112220510	AFG3L1P: Intron variant	WT/R	-5.90	1.16	**2.57E-05**
		R/R	4.16	4.46	0.359032
rs59713793	NA	WT/R	-5.79	1.73	**0.002522**
		R/R	-2.37	1.75	0.187888
rs34248444	DBNDD1: Downstream variant	WT/R	2.18	1.64	0.193496
rs112388869	TUBB3: Intron variant	WT/R	-4.07	2.01	0.052955

Variant	Gene: Consequence	Genotype	β	SD	p-value
rs11446223	DEF8: Intrion variant	WT/R	6.06	2.18	**0.009843**
		R/R	11.62	3.57	**0.003125**
rs11453006	DEF8: Intron variant	WT/R	-7.88	2.09	**0.000842**
		R/R	-7.50	3.22	**0.028156**

WT - Wild type allele; R - Effect allele; β - variation in Δage in relation to the reference genotype (most common genotype); SD - standard deviation; NA - Non available.

Supplemental Information

Table S1. Extracted variants from VCF files

Locus	Gene	CHR	MBP	Type	Alleles
rs10584116	AFG3L1P	16	90.0	Indel	delCT
rs5818729	AFG3L1P	16	90.0	Indel	delCC
rs112220510	AFG3L1P	16	90.0	Indel	delA
rs112373300	AFG3L1P	16	90.0	Indel	insTTAA
rs112556696	AFG3L1P	16	90.0	SNP	A>G
rs113020559	AFG3L1P	16	90.0	Indel	delA
rs11365264	AFG3L1P	16	90.0	Indel	delA
rs138606117	AFG3L1P	16	90.0	Indel	delG
rs140607816	AFG3L1P	16	90.0	Indel	delG
rs147227691	AFG3L1P	16	90.0	Indel	delC
rs147810494	AFG3L1P	16	90.0	Indel	delA
rs148467702	AFG3L1P	16	90.0	Indel	delTT
rs149262318	AFG3L1P	16	90.0	Indel	delCACAAAAAA
rs151289005	AFG3L1P	16	90.0	Indel	insA
rs60409378	AFG3L1P	16	90.0	Indel	insTA
rs199824666	AFG3L1P	16	90.0	Indel	delG
rs200234556	AFG3L1P	16	90.0	Indel	delGATC
rs201156703	AFG3L1P	16	90.0	Indel	delTA
rs34239467	AFG3L1P	16	90.0	Indel	delC
rs34850152	AFG3L1P	16	90.0	Indel	insG
rs36089706	AFG3L1P	16	90.0	Indel	delT
rs5818730	AFG3L1P	16	90.0	Indel	delA
rs5818731	AFG3L1P	16	90.0	Indel	delT
rs80032725	AFG3L1P	16	90.0	Indel	delGCCACCTACTCCT
rs34248444	DBNDD1	16	90.0	Indel	delA
rs10573313	DEF8	16	90.0	Indel	delTGGT
rs111421479	DEF8	16	90.0	Indel	delT
rs112881095	DEF8	16	90.0	Indel	delT
rs113500176	DEF8	16	90.0	Indel	delCT
rs11446223	DEF8	16	90.0	Indel	delA
rs11453006	DEF8	16	90.0	Indel	delT
rs149156863	DEF8	16	90.0	Indel	insA
rs370010270	DEF8	16	90.0	Indel	delT
rs56857645	DEF8	16	90.0	Indel	delG
rs12931267	FANCA	16	89.8	SNP	C>G
rs75570604	FANCA	16	89.8	SNP	G>C
rs1805005	MC1R	16	89.9	SNP	G>T
rs1805007	MC1R	16	89.9	SNP	C>T
rs1805008	MC1R	16	89.9	SNP	C>T

Locus	Gene	CHR	MBP	Type	Alleles
rs1805009	MC1R	16	89.9	SNP	G>C
rs312262906	MC1R	16	89.9	Indel	insA
rs10611547	NA	NA	NA	Indel	delAGGTGGAGG
rs11276058	NA	NA	NA	Indel	delGATGACTTAGTACTGCCCCTCAGCAGG
rs149945057	NA	16	90.0	Indel	insG
rs4785704	NA	16	89.7	SNP	A>G
rs58851373	NA	16	90.0	Indel	delA
rs59713793	NA	16	90.0	Indel	delCAG
rs72468543	NA	NA	NA	Indel	delCTCAGCAGGGATGA
rs34265416	SPATA2L	16	89.7	SNP	C>A
rs36100920	SPIRE2	16	89.9	Indel	delAA
rs112388869	TUBB3	16	89.9	Indel	insC
rs150764400	TUBB3	16	89.9	Indel	delC
rs201112914	TUBB3	16	89.9	Indel	insT
rs59751572	TUBB3	16	89.9	Indel	delCCTTCCTT
rs372052072	TUBB3	16	89.9	Indel	delC
rs57346585	TUBB3	16	89.9	Indel	insTCT
rs34714188	VPS9D1	16	89.7	SNP	T>A
rs12924124	ZNF276	16	89.7	SNP	C>T
rs35026726	ZNF276	16	89.7	SNP	C>T

CHR - Chromossome; MBP - Mega base pair position; NA - Non available.

Table S2. Variant's population genetics indices and forensic parameters

Locus	A	C	G	T	WT	R	GD	PIC	PM	PD	Hobs	PE	TPI	pHW
rs10573313					0.998	0.002	0.004	0.004	0.992	0.008	0.004	1.46E+08	0.502	1.000
rs10584116					0.833	0.167	0.278	0.239	0.566	0.434	0.249	0.045	0.666	0.129
rs5818729					0.006	0.994	0.011	0.011	0.977	0.023	0.011	0.000	0.506	1.000
rs10611547					0.107	0.893	0.192	0.173	0.672	0.328	0.192	0.028	0.618	1.000
rs111421479					0.807	0.193	0.313	0.263	0.522	0.478	0.318	0.071	0.733	0.838
rs112220510					0.791	0.209	0.331	0.276	0.504	0.496	0.303	0.065	0.717	0.188
rs112373300					0.092	0.908	0.167	0.153	0.707	0.293	0.169	0.022	0.601	1.000
rs112388869					0.969	0.031	0.060	0.058	0.885	0.115	0.061	0.003	0.533	1.000
rs112556696	0.858		0.142				0.244	0.214	0.614	0.386	0.207	0.032	0.630	0.022
rs11276058					0.948	0.052	0.098	0.093	0.820	0.180	0.096	0.008	0.553	0.516
rs112881095					0.989	0.011	0.023	0.022	0.955	0.045	0.023	0.001	0.512	1.000
rs113020559					0.856	0.144	0.247	0.216	0.600	0.400	0.241	0.042	0.659	0.788
rs113500176					0.856	0.144	0.247	0.216	0.600	0.400	0.241	0.042	0.659	0.806
rs11365264					0.425	0.575	0.490	0.369	0.393	0.607	0.513	0.200	1.028	0.447
rs11446223					0.805	0.195	0.315	0.265	0.521	0.479	0.299	0.063	0.713	0.431
rs11453006					0.812	0.188	0.306	0.258	0.532	0.468	0.284	0.057	0.698	0.307
rs12924124		0.895		0.105			0.189	0.171	0.700	0.300	0.142	0.016	0.583	0.001
rs12931267		0.898	0.102				0.183	0.166	0.706	0.294	0.142	0.016	0.583	0.002
rs138606117					0.992	0.008	0.015	0.015	0.970	0.030	0.015	0.000	0.508	1.000

rs140607816					0.998	0.002	0.004	0.004	0.992	0.008	0.004	1.46E+08	0.502	1.000
rs147227691					0.989	0.011	0.023	0.022	0.955	0.045	0.023	0.001	0.512	1.000
rs147810494					0.992	0.008	0.015	0.015	0.970	0.030	0.015	0.000	0.508	1.000
rs148467702					0.990	0.010	0.019	0.019	0.970	0.030	0.011	0.000	0.506	0.021
rs149156863					0.994	0.006	0.011	0.011	0.977	0.023	0.011	0.000	0.506	1.000
rs149262318					0.998	0.002	0.004	0.004	0.992	0.008	0.004	1.46E+08	0.502	1.000
rs149945057					0.994	0.006	0.011	0.011	0.977	0.023	0.011	0.000	0.506	1.000
rs150764400					0.977	0.023	0.045	0.044	0.912	0.088	0.046	0.002	0.524	1.000
rs151289005					0.987	0.013	0.027	0.026	0.948	0.052	0.027	0.001	0.514	1.000
rs1805005			0.935	0.065			0.122	0.114	0.788	0.212	0.107	0.010	0.560	0.090
rs1805007		0.895		0.105			0.189	0.171	0.691	0.309	0.157	0.019	0.593	0.014
rs1805008		0.931		0.069			0.129	0.120	0.762	0.238	0.138	0.015	0.580	0.138
rs1805009		0.025	0.975				0.049	0.047	0.905	0.095	0.050	0.002	0.526	1.000
rs60409378					0.839	0.161	0.271	0.234	0.571	0.429	0.261	0.049	0.676	0.506
rs199824666					0.998	0.002	0.004	0.004	0.992	0.008	0.004	1.46E+08	0.502	1.000
rs200234556					0.994	0.006	0.011	0.011	0.977	0.023	0.011	0.000	0.506	1.000
rs201112914					0.864	0.136	0.235	0.207	0.609	0.391	0.249	0.045	0.666	0.432
rs201156703					0.897	0.103	0.186	0.168	0.692	0.308	0.161	0.020	0.596	0.044
rs312262906					0.996	0.004	0.008	0.008	0.985	0.015	0.008	5.78E+09	0.504	1.000
rs34239467					0.243	0.757	0.369	0.300	0.472	0.528	0.402	0.115	0.837	0.176
rs34248444					0.985	0.015	0.030	0.030	0.941	0.059	0.031	0.001	0.516	1.000
rs34265416	0.123	0.877					0.216	0.192	0.655	0.345	0.176	0.024	0.607	0.006
rs34714188	0.103			0.897			0.186	0.168	0.705	0.295	0.138	0.015	0.580	0.000
rs34850152					0.971	0.029	0.056	0.054	0.892	0.108	0.057	0.003	0.530	1.000
rs35026726		0.895		0.105			0.189	0.171	0.700	0.300	0.142	0.016	0.583	0.000
rs36089706					0.920	0.080	0.148	0.137	0.738	0.262	0.146	0.017	0.585	0.684
rs36100920					0.109	0.891	0.195	0.176	0.694	0.306	0.142	0.016	0.583	0.001
rs59751572					0.226	0.774	0.351	0.289	0.485	0.515	0.307	0.066	0.721	0.046
rs372052072					0.877	0.123	0.216	0.192	0.645	0.355	0.199	0.030	0.624	0.262
rs4785704	0.877		0.123				0.216	0.192	0.655	0.345	0.176	0.024	0.607	0.005
rs56857645					0.992	0.008	0.015	0.015	0.970	0.030	0.015	0.000	0.508	1.000
rs57346585					0.678	0.322	0.437	0.341	0.434	0.566	0.490	0.179	0.981	0.073
rs5818730					0.427	0.573	0.490	0.370	0.395	0.605	0.517	0.203	1.036	0.380
rs5818731					0.427	0.573	0.490	0.370	0.395	0.605	0.517	0.203	1.036	0.395
rs58851373					0.810	0.190	0.308	0.260	0.533	0.467	0.264	0.050	0.680	0.030
rs59713793					0.852	0.148	0.252	0.220	0.646	0.354	0.126	0.013	0.572	0.000
rs72468543					0.893	0.107	0.192	0.173	0.669	0.331	0.199	0.030	0.624	0.745
rs75570604		0.107	0.893				0.192	0.173	0.690	0.310	0.153	0.018	0.590	0.004
rs80032725					0.971	0.029	0.056	0.054	0.898	0.102	0.050	0.002	0.526	1.000

A/C/T/G - Alleles frequencies; WT - Wild type allele frequency; R - Effect allele frequency; GD - Gene Diversity; PIC -Polymorphism information content; PM - Random match probability; PD - Power of discrimination; Hobs - Observed heterozigosity; PE - Power of exclusion; TPI - Typical paternity index; pHW - p-value of a teset of Hardy-Weinberg Equilibrium (HWE). A low p-value indicates a significant deviation from HWE.

Appendix A - Data collection form model

FORENSIC DNA PHENOTYPING - BRAZIL | INSTRUMENTO DE DADOS

DATA DA COLETA:_____ / _____ / _____ **IDENTIFICAÇÃO:** ☐☐☐☐☐☐☐☐

DADOS PESSOAIS DO PARTICIPANTE

Nome: _____ Data de Nascimento: _____ / _____ / _____

Contato (E-mail; telefones): _____

DADOS ANTROPOMÉTRICOS / CARACTERÍSTICAS / INFORMAÇÕES

Altura: _____ Peso: _____ N^0 do sapato: _____

Ao usar a mão para escrever é:	( ) Destro	( ) Canhoto			
A orelha é saliente?	( ) SIM	( ) NÃO	Lóbulo orelha é solto?	( ) SIM	( ) NÃO
Está usando lente de contato?	( ) SIM	( ) NÃO	Está bem bronzeado?	( ) SIM	( ) NÃO
Cabelo pintado/descolorido...?	( ) SIM	( ) NÃO	...Alisado/Encrespado?	( ) SIM	( ) NÃO

DADOS EPIDEMIOLÓGICOS

Sexo:	( ) Masc	( ) Fem				
Origem:	( ) Europeia	( ) Africana	( ) Indígena	( ) Asiática	( ) Árabe*	( ) Outra: _____
Residente em:	( ) Sul	( ) Sudeste	( ) Centro-Oeste	( )Nordeste	( ) Norte	Estado/Cidade:
Natural de:	( ) Sul	( ) Sudeste	( ) Centro-Oeste	( )Nordeste	( ) Norte	Estado/Cidade:

(***Árabes**: Habitantes da Península Arábica, Oriente Médio, África setentrional).

DADOS BIOLÓGICOS - ORIGINAIS (NATURAIS)

OLHO COR:	( ) Azul Claro	( ) Azul Esc.	( ) Verde	( ) Mel	( ) Cast. Claro	( ) Cast. Esc.	( ) Preto
PELE COR:	( ) Tipo I	( ) Tipo II	( ) Tipo III	( ) Tipo IV		( ) Tipo V	( ) Tipo VI
CABELO COR:	( ) Loiro Plat.	( ) Loiro Amar.	( ) Loiro Esc.	( ) Castanho Claro		( ) Cast. Esc.	( ) Negro
	( ) Ruivo Claro	( ) Ruivo Esc.					
CABELO TIPO:	( ) Liso	( ) Ondulado	( ) Crespo	( ) Molinha			
CABELO QTDE:	( ) Muito	( ) Médio	( ) Pouco	**CABELO OBSERV.**:		( ) Calvície	( ) Grisalho
						( ) Branco	( ) _____

PARA HOMENS

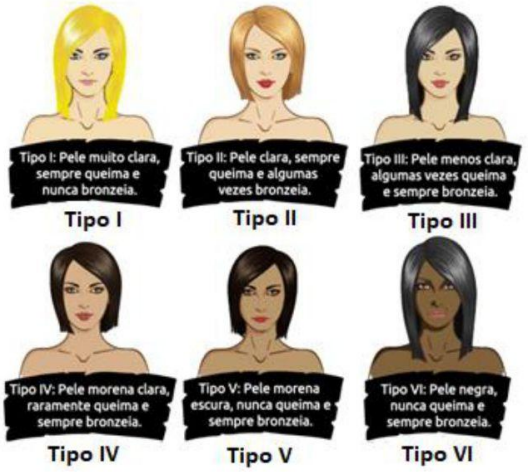
QTD DE PELO EM FACE/PEITO/COSTAS/PERNAS/BRAÇOS: ( ) Muito ( ) Médio ( ) Pouco

COLETADOR

Nome COLETADOR:		**Matrícula PUCRS:**	
Como coletador, declaro que colhi e identifiquei corretamente todas as amostras de células da pessoa aqui indicada para fins de investigação de dados fenotípicos, tendo seguido rigorosamente os procedimentos indicados. Para tanto assino ao lado:		_____ **Assinatura COLETADOR**	

OBSERVAÇÕES

(SE FOR O CASO: DESENHE O HEREDOGRAMA DA FAMÍLIA À QUAL O PARTICIPANTE PERTENCE, IDENTIFICANDO-O COM UMA SETA).



Appendix B - Consent form model

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO (TALE) NO DESENVOLVIMENTO DA PRESENTE PESQUISA REFERENTE AO ESTUDO DE FENÓTIPO (POPULAÇÃO)

Você está sendo convidado para participar de uma pesquisa MARCADORES GENÉTICOS PARA AS CARACTERÍSTICAS EXTERNAS VISÍVEIS.

Nós comparamos o material genético, cor dos olhos/pele/cabelo, formato do rosto, peso e altura das pessoas, para saber se estão associados.

Nós precisamos: 1- anotar seus dados
2- fazer umas 10-12 fotos suas
3- passar um cotonete dentro da sua boca

O adulto responsável por você assinará um TERMO (TCLE) permitindo sua participação, mas você pode expressar seu próprio desejo por meio deste documento. Você pode fazer qualquer pergunta sobre isso, e também pode preferir não fazer.

Eu, (preencher com nome do participante) fui informado dos objetivos da pesquisa e entendi, e concordo em participar deste estudo. Por isso, deixo aqui meu consentimento. Se tiver novas perguntas sobre a pesquisa, posso telefonar para a Dra. Clarice S. Alho no telefone (51) 3320 3545, ramal 4534, ou para esclarecer sobre os meus direitos como participante deste estudo, ou se penso que fui prejudicado pela minha participação, posso chamar também os responsáveis pelo Comitê de Ética em Pesquisa da PUCRS (CEP-PUCRS) no telefone (51) 3320 3000 ramal 3345. Entendi que o material e os dados obtidos a partir desta pesquisa poderão ser usados em pesquisas futuras, feitas aqui ou fora do país, mas isso acontecerá se elas também forem avaliadas e aprovadas pelo CEP/CONEP. Se uma nova pesquisa aprovada pelo CEP/CONEP for realizada, eu:

( ) concordo que podem usar esse mesmo termo.
( ) prefiro assinar um novo termo.

E-mail / telefones: _____

Nome e assinatura do participante	**Local e data**
Nome e assinatura do pesquisador	**Local e data**

- ✓ O CEP-PUCRS tem como endereço: Avenida Ipiranga, 6690 - Prédio 40 - 5° andar – Sala 505 – Porto Alegre/RS; telefone (51) 3320 3000 ramal 3345.
- ✓ Este Termo de Consentimento foi elaborado em DUAS VIAS, sendo uma ficará com você e a outra com o pesquisador, conforme a Resolução CNS 466/2012 itens IV.3.f, IV.5.d.

5 CONCLUSÕES E PERSPECTIVAS

No presente estudo, fomos capazes de demonstrar, pela primeira vez, a influência de polimorfismos SNPs e INDEL presentes em regiões intrônicas de genes adjacentes ao locus *MC1R* - como *TUBB3* rs59751572*, FANCA* rs12931267*, AFG3L1P* rs201156703 e *ZNF276* rs12924124 - e a aparência facial etária de indivíduos da população do Rio Grande do Sul. As variantes mostraram influenciar tanto para fenótipos de face mais jovens quanto para fenótipos de face mais velhos, diferente do observado em populações europeias no estudo de Liu et al. (2016), que serviu como base para a realização deste trabalho. Além disso, como a influência destas variantes na PA parece bastante drástica e estar sendo superestimada, é importante que elas sejam consideradas em conjunto no contexto de nosso modelo estatístico para que seu efeito esteja balanceado. Isso é justificável baseado no fato de que determinadas variantes estejam correlacionadas, indicando uma segregação não-independente na formação dos haplótipos. Ainda que os mecanismos moleculares por detrás do efeito das variantes na PA não terem sido elucidados até o presente momento, propomos a hipótese de que elas possam estar interferindo em potenciais regiões regulatórias do gene *MC1R,* assim como em sítios de *splicing* intergênicos.

Apesar de nossos resultados, alguns pontos são válidos de serem mencionados como possíveis limitadores do trabalho. Nosso instrumento de coleta de dados dos participantes não inclui questionamentos acerca de informações relevantes dos participantes que podem influenciar em sua PA (*e.g.:* tabagismo, atividade laboral envolvendo exposição demasiada à radiação UV, presença de intervenções estéticas na face, uso de medicamentos que interferem no processo de melanogênese, etc.). Ainda, visto que as fotografias frontais de face, mesmo que padronizadas, são originárias de documentos de identificação oficial, não foi possível realizar o controle de elementos que pudessem ter influência semelhante, como uso de maquiagens, excesso de pelos faciais ou pequenas mudanças de expressão. A presença de marcas d'água que não puderam ser removidas também impediu a utilização de modelos *deep learning* para heteroclassificação etária, etapa que originalmente constava no projeto e que poderia ter enriquecido nossos resultados. Apesar de estes pontos não terem sido cobertos no estudo, eles não refletem a realidade encontrada durante um processo de investigação, e de forma alguma invalidam nossos resultados.

Por fim, vale ressaltar que, mesmo que seja importante estabelecer potenciais marcadores moleculares capazes de prever *EVCs* como a PA em populações miscigenadas, os resultados fornecidos pela aplicação da *FDP* na prática forense são probabilísticos, e não determinísticos. Como já mencionado em seções anteriores deste trabalho, *EVCs* são em sua maioria traços complexos, o que confere aos fenótipos uma margem de variação.

REFERÊNCIAS

ABDEL-MALEK, Zalfa A. *et al.* α-MSH tripeptide analogs activate the melanocortin 1 receptor and reduce UV-induced DNA damage in human melanocytes. **Pigment Cell & Melanoma Research**, [*s. l.*], v. 22, n. 5, p. 635–644, 2009.

ARNOLD, Carrie. The controversial company using DNA to sketch the faces of criminals. **Nature**, [*s. l.*], v. 585, n. 7824, p. 178–181, 2020.

BRESLIN, Krystal *et al.* HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms. **Forensic Science International: Genetics**, [*s. l.*], v. 43, 2019. Disponível em: https://www.fsigenetics.com/article/S1872-4973(19)30205-4/fulltext. Acesso em: 8 jun. 2022.

BRITO, Ana; PONTES, A.N. ESTUDO SOBRE IDENTIFICAÇÃO HUMANA POR DNA, BANCO NACIONAL DE PERFIS GENÉTICOS E A OBRIGATORIEDADE DA LEI 12654/2012. **Revista Brasileira de Criminalística**, [*s. l.*], v. 9, p. 75, 2020.

BUDOWLE, Bruce; VAN DAAL, Angela. Forensically relevant SNP classes. **BioTechniques**, [*s. l.*], v. 44, n. 5, p. 603–608, 610, 2008.

BUTLER, John M.; COBLE, Michael D.; VALLONE, Peter M. STRs vs. SNPs: thoughts on the future of forensic DNA testing. **Forensic Science, Medicine, and Pathology**, [*s. l.*], v. 3, n. 3, p. 200–205, 2007.

CANALES SERRANO, Aurora. Forensic DNA phenotyping: A promising tool to aid forensic investigation. Current situation. **Spanish Journal of Legal Medicine**, [*s. l.*], v. 46, n. 4, p. 183–190, 2020.

CERQUEIRA, Caio Cesar Silva de. **Relação genótipo a fenótipo e a pigmentação humana: aspectos evolutivos e sua aplicação na genética forense.** 2013. 308 f. Tese de doutorado apresentada no Programa de Pós-graduação em Genética e Biologia Molecular - Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2013.

DIAS, Helena Correia *et al.* DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay. **Forensic Science International**, [*s. l.*], v. 311, p. 110267, 2020.

DIAS FILHO, Claudemir Rodrigues *et al.* **Introdução à genética forense**. Campinas, SP.: Millenium Editora Ltda., 2020. (Ciência contra o crime).

DNA FORENSICS — PARABON NANOLABS. [*S. l.*], [*s. d.*]. Disponível em: https://parabon-nanolabs.com/dna-forensics/index.html. Acesso em: 8 jun. 2022.

FRANCEZ, Pablo; LIMA, Adriano; ALMEIDA, Rivelton. Relationship between Ancestry Inferred By Molecular Analysis, Self-Report and Hetero-Classification. **Journal of Forensic Research**, [*s. l.*], v. 06, 2015.

GARCÍA-BORRÓN, Jose C; ABDEL-MALEK, Zalfa; JIMÉNEZ-CERVANTES, Celia. MC1R, the cAMP pathway and the response to solar UV: Extending the horizon beyond pigmentation. **Pigment cell & melanoma research**, [*s. l.*], v. 27, n. 5, p. 699–720, 2014.

GUNN, David A. *et al.* Why Some Women Look Young for Their Age. **PLoS ONE**, [*s. l.*], v. 4, n. 12, p. e8021, 2009.

HONG, Sae Rom *et al.* DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. **Forensic Science International: Genetics**, [*s. l.*], v. 29, p. 118–125, 2017.

HORRELL, Erin; BOULANGER, Mary; D'ORAZIO, John. Melanocortin 1 Receptor: Structure, Function, and Regulation. **Frontiers in Genetics**, [*s. l.*], v. 7, 2016.

JUNG, Sang-Eun *et al.* DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. **Forensic Science International. Genetics**, [*s. l.*], v. 38, p. 1–8, 2019.

KADEKARO, Ana Luisa *et al.* α-Melanocortin and Endothelin-1 Activate Antiapoptotic Pathways and Reduce DNA Damage in Human Melanocytes. **Cancer Research**, [*s. l.*], v. 65, n. 10, p. 4292–4299, 2005.

KAYSER, Manfred. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. **Forensic Science International. Genetics**, [*s. l.*], v. 18, p. 33–48, 2015.

KAYSER, Manfred; SCHNEIDER, Peter M. DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. **Forensic Science International: Genetics**, [*s. l.*], v. 3, n. 3, p. 154–161, 2009.

KOOPS, Bert-Jaap; SCHELLEKENS, Maurice H. M. **Forensic DNA Phenotyping: Regulatory Issues**. Rochester, NY: Social Science Research Network, 2006. SSRN Scholarly Paper. Disponível em: https://papers.ssrn.com/abstract=975032. Acesso em: 8 jun. 2022.

KOWALCZYK, Marek *et al.* Molecular markers used in forensic genetics. **Medicine, Science, and the Law**, [*s. l.*], v. 58, n. 4, p. 201–209, 2018.

LANDER, Eric S. *et al.* Initial sequencing and analysis of the human genome. **Nature**, [*s. l.*], v. 409, n. 6822, p. 860–921, 2001.

LAW, Matthew H. *et al.* Genome-Wide Association Shows that Pigmentation Genes Play a Role in Skin Aging. **Journal of Investigative Dermatology**, [*s. l.*], v. 137, n. 9, p. 1887–1894, 2017.

LIU, Fan *et al.* The MC1R Gene and Youthful Looks. **Current biology: CB**, [*s. l.*], v. 26, n. 9, p. 1213–1220, 2016.

MULLANEY, Julienne M. *et al.* Small insertions and deletions (INDELs) in human genomes. **Human Molecular Genetics**, [*s. l.*], v. 19, n. R2, p. R131–R136, 2010.

NKENGNE, A *et al.* Influence of facial skin attributes on the perceived age of Caucasian women. **Journal of the European Academy of Dermatology and Venereology**, [*s. l.*], v. 22, n. 8, p. 982–991, 2008.

NUSSBAUM, Robert L.; MCINNES, Roderick R.; WILLARD, Huntington F. **Thompson & Thompson, genética médica**. 7. ed. Rio de Janeiro: Elsevier Editora Ltda., 2008.

PENA, S. D. J. *et al.* DNA tests probe the genomic ancestry of Brazilians. **Brazilian Journal of Medical and Biological Research**, [*s. l.*], v. 42, p. 870–876, 2009.

POWERPLEX® FUSION 6C SYSTEM. [*S. l.*], [*s. d.*]. Disponível em: https://www.promega.com.br/products/forensic-dna-analysis-ce/str-amplification/powerplex-fusion-6c-system/. Acesso em: 10 out. 2021.

PRAINSACK, Barbara; SAMUEL, Gabrielle. The regulatory landscape of forensic DNA phenotyping in Europe VISAGE. [*s. l.*], 2018.

RUIZ-LINARES, Andrés *et al.* Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. **PLoS Genetics**, [*s. l.*], v. 10, n. 9, p. e1004572, 2014.

SALOUM DE NEVES MANTA, Fernanda *et al.* Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. **PLoS ONE**, [*s. l.*], v. 8, n. 9, p. e75145, 2013.

SCHNEIDER, Peter M.; PRAINSACK, Barbara; KAYSER, Manfred. The Use of Forensic DNA Phenotyping in Predicting Appearance and Biogeographic Ancestry. **Deutsches Arzteblatt International**, [*s. l.*], v. 51–52, n. 51–52, p. 873–880, 2019.

SOUZA, Aracele Maria de *et al.* A systematic scoping review of the genetic ancestry of the Brazilian population. **Genetics and Molecular Biology**, [*s. l.*], v. 42, p. 495–508, 2019.

STRANGER, Barbara E.; STAHL, Eli A.; RAJ, Towfique. Progress and promise of genome-wide association studies for human complex trait genetics. **Genetics**, [*s. l.*], v. 187, n. 2, p. 367–383, 2011.

VALVERDE, Paloma *et al.* Variants of the melanocyte–stimulating hormone receptor gene are associated with red hair and fair skin in humans. **Nature Genetics**, [*s. l.*], v. 11, n. 3, p. 328–330, 1995.

VIDAKI, Athina; KAYSER, Manfred. Recent progress, methods and perspectives in forensic epigenetics. **Forensic Science International. Genetics**, [*s. l.*], v. 37, p. 180–195, 2018.

VIRMOND, Marina *et al.* Fenotipagem forense pelo DNA através de SNPs. **Revista Brasileira de Criminalística**, [*s. l.*], v. 5, p. 37, 2016.

VOEGELI, Rainer *et al.* Differences between perceived age and chronological age in women: A multi-ethnic and multi-centre study. **International Journal of Cosmetic Science**, [*s. l.*], v. 43, n. 5, p. 547–560, 2021.

WALSH, Susan *et al.* Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. **Forensic Science International: Genetics**, [*s. l.*], v. 5, n. 5, p. 464–471, 2011.

WALSH, Susan *et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. **Forensic Science International: Genetics**, [*s. l.*], v. 7, n. 1, p. 98–115, 2013.

WEBER, James *et al.* Human Diallelic Insertion/Deletion Polymorphisms. **American journal of human genetics**, [*s. l.*], v. 71, p. 854–862, 2002.

WYNER, Nicole; BARASH, Mark; MCNEVIN, Dennis. Forensic Autosomal Short Tandem Repeats and Their Potential Association With Phenotype. **Frontiers in Genetics**, [*s. l.*], v. 11, 2020. Disponível em: https://www.frontiersin.org/articles/10.3389/fgene.2020.00884/full. Acesso em: 28 abr. 2021.

ZUBAKOV, Dmitry *et al.* Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. **Forensic Science International. Genetics**, [*s. l.*], v. 24, p. 33–43, 2016.

ANEXO A - NORMAS DE PUBLICAÇÃO DA REVISTA FORENSIC SCIENCE INTERNATIONAL: GENETICS

Author Information

Your Paper Your Way

We now differentiate between the requirements for new and revised submissions. You may choose to submit your manuscript as a single Word or PDF file to be used in the refereeing process. Only when your paper is at the revision stage, will you be requested to put your paper in to a 'correct format' for acceptance and provide the items required for the publication of your article.
To find out more, please visit the Preparation section below.

Introduction

Types of paper

1. Original Research Papers (Regular Papers)
2. Review Articles
3. Short Communications
4. Book Reviews
5. Case Reports
6. Correspondence
7. Forensic Population Genetics (Original Paper, Short Communication or Correspondence)

Case Reports will be accepted only if they contain some important new information for the readers.

Forensic Population Genetics manuscripts can be submitted using three types of formats:

Forensic Population Genetics - Original paper: In this section full length papers on relevant population genetics issues of forensic interest will be considered for publication. The data should be original, the population genetic analysis must be of the highest quality and the data should have

forensic relevance beyond the scope of simply reporting allele or haplotype frequencies.

Forensic Population Genetics - Short communication (former "Announcement of population data"): Understanding that both the quality of population data and the relevance of results are crucial short communications should be submitted in table format. Population data are required to be downloaded as supplementary files (see Preparation of supplementary data).

Forensic Population Genetics - Correspondence: If the relevance of the data is not sufficient for an original paper or a short communication, but still worthy of an announcement, the editors can invite authors to submit a letter to the editor. In this case the manuscript must be written in the form of a short letter to the editor summarizing the relevant information while the frequency data must be provided as an electronic supplement, e.g. a spreadsheet table, for online publication in the electronic repository of the journal.

All Forensic Population Genetics papers should always contain information on the description of the population, ethical requirements and quality control. For mtDNA DNA papers, previous acceptance of the dataset in EMPOP (http://www.empop.org) is required, for YSTR and YSNP data previous inclusion of the data in the YSTR/YSNP database (http://www.yhrd.org) is required. For specific information on requirements and procedures of Forensic Population Genetics papers, see the latest ethical requirements for publication: https://www.sciencedirect.com/science/article/pii/S1872497320300727.
Failure to adhere to this guidance will result in an automatic rejection. The editor?s decision is final in this regard.

Rapid Communications should describe work of significant interest, whose impact would suffer if publication were not expedited. They should not be longer than 5 printed journal pages. Authors may suggest that their work is treated as a Rapid Communication, but the final decision on whether it is suitable as such will be taken by the Editor. Rapid Communications requiring revision should be resubmitted as a new submission.

Please note: Any articles which refer to FSIGEN or ISFG guidelines must include these in the body of the text and not as a citation.

Contact details for submission

For queries concerning the submission process or journal procedures please visit the Elsevier Support Center. Authors can check the status of their manuscript within the review procedure using Elsevier Editorial System.

Submission checklist

You can use this list to carry out a final check of your submission before you send it to the journal for review. Please check the relevant section in this Guide for Authors for more details.

Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:
• E-mail address
• Full postal address

All necessary files have been uploaded:
Manuscript:
• Include keywords
• All figures (include relevant captions)
• All tables (including titles, description, footnotes)
• Ensure all figure and table citations in the text match the files provided
• Indicate clearly if color should be used for any figures in print
Graphical Abstracts / Highlights files (where applicable)
Supplemental files (where applicable)

Further considerations
• Manuscript has been 'spell checked' and 'grammar checked'
• All references mentioned in the Reference List are cited in the text, and vice versa
• Permission has been obtained for use of copyrighted material from other sources (including the Internet)

• A competing interests statement is provided, even if the authors have no competing interests to declare
• Journal policies detailed in this guide have been reviewed
• Referee suggestions and contact details provided, based on journal requirements

For further information, visit our Support Center.

Before You Begin

Ethics in publishing

Please see our information on Ethics in publishing.

Studies in humans and animals

If the work involves the use of human subjects, the author should ensure that the work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The manuscript should be in line with the Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals and aim for the inclusion of representative human populations (sex, age and ethnicity) as per those recommendations. The terms sex and gender should be used correctly.

Authors should include a statement in the manuscript that informed consent was obtained for experimentation with human subjects. The privacy rights of human subjects must always be observed.

All animal experiments should comply with the ARRIVE guidelines and should be carried out in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and associated guidelines, EU Directive 2010/63/EU for animal experiments, or the National Research Council's Guide for the Care and Use of Laboratory Animals and the authors should clearly indicate in the manuscript that such guidelines have been followed. The sex of animals must be indicated, and where appropriate, the influence (or association) of sex on the results of the study.

Declaration of competing interest

All authors must disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work. Examples of potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding. Authors should complete the declaration of competing interest statement using this template and upload to the submission system at the Attach/Upload Files step. Note: Please do not convert the .docx template to another file type. Author signatures are not required. If there are no interests to declare, please choose the first option in the template. More information.

Submission declaration and verification

Submission of an article implies that the work described has not been published previously (except in the form of an abstract, a published lecture or academic thesis, see 'Multiple, redundant or concurrent publication' for more information), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. To verify originality, your article may be checked by the originality detection service Crossref Similarity Check.

Preprints

Please note that preprints can be shared anywhere at any time, in line with Elsevier's sharing policy. Sharing your preprints e.g. on a preprint server will not count as prior publication (see 'Multiple, redundant or concurrent publication' for more information).

Use of inclusive language

Inclusive language acknowledges diversity, conveys respect to all people, is sensitive to differences, and promotes equal opportunities. Content should make no assumptions about the beliefs or commitments of any reader; contain nothing which might imply that one individual is superior to another on the grounds of age, gender, race, ethnicity, culture, sexual orientation,

disability or health condition; and use inclusive language throughout. Authors should ensure that writing is free from bias, stereotypes, slang, reference to dominant culture and/or cultural assumptions. We advise to seek gender neutrality by using plural nouns ("clinicians, patients/clients") as default/wherever possible to avoid using "he, she," or "he/she." We recommend avoiding the use of descriptors that refer to personal attributes such as age, gender, race, ethnicity, culture, sexual orientation, disability or health condition unless they are relevant and valid. When coding terminology is used, we recommend to avoid offensive or exclusionary terms such as "master", "slave", "blacklist" and "whitelist". We suggest using alternatives that are more appropriate and (self-) explanatory such as "primary", "secondary", "blocklist" and "allowlist". These guidelines are meant as a point of reference to help identify appropriate language but are by no means exhaustive or definitive.

Author contributions

For transparency, we encourage authors to submit an author statement file outlining their individual contributions to the paper using the relevant CRediT roles: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft; Writing - review & editing. Authorship statements should be formatted with the names of authors first and CRediT role(s) following. More details and an example.

Changes to authorship

Authors are expected to consider carefully the list and order of authors before submitting their manuscript and provide the definitive list of authors at the time of the original submission. Any addition, deletion or rearrangement of author names in the authorship list should be made only before the manuscript has been accepted and only if approved by the journal Editor. To request such a change, the Editor must receive the following from the corresponding author: (a) the reason for the change in author list and (b) written confirmation (e-mail, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed.

Only in exceptional circumstances will the Editor consider the addition, deletion or rearrangement of authors after the manuscript has been accepted. While the Editor considers the request, publication of the manuscript will be suspended. If the manuscript has already been published in an online issue, any requests approved by the Editor will result in a corrigendum.

Article transfer service

This journal is part of our Article Transfer Service. This means that if the Editor feels your article is more suitable in one of our other participating journals, then you may be asked to consider transferring the article to one of those. If you agree, your article will be transferred automatically on your behalf with no need to reformat. Please note that your article will be reviewed again by the new journal. More information.

Copyright

Upon acceptance of an article, authors will be asked to complete a 'Journal Publishing Agreement' (see more information on this). An e-mail will be sent to the corresponding author confirming receipt of the manuscript together with a 'Journal Publishing Agreement' form or a link to the online version of this agreement.

Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution and for all other derivative works, including compilations and translations. If excerpts from other copyrighted works are included, the author(s) must obtain written permission from the copyright owners and credit the source(s) in the article. Elsevier has preprinted forms for use by authors in these cases.

For gold open access articles: Upon acceptance of an article, authors will be asked to complete a 'License Agreement' (more information). Permitted third party reuse of gold open access articles is determined by the author's choice of user license.

Author rights
As an author you (or your employer or institution) have certain rights to reuse your work. More information.

Elsevier supports responsible sharing

Find out how you can share your research published in Elsevier journals.

Role of the funding source

You are requested to identify who provided financial support for the conduct of the research and/or preparation of the article and to briefly describe the role of the sponsor(s), if any, in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. If the funding source(s) had no such involvement, it is recommended to state this.

Open access

Please visit our Open Access page for more information.

Elsevier Researcher Academy

Researcher Academy is a free e-learning platform designed to support early and mid-career researchers throughout their research journey. The "Learn" environment at Researcher Academy offers several interactive modules, webinars, downloadable guides and resources to guide you through the process of writing for research and going through peer review. Feel free to use these free resources to improve your submission and navigate the publication process with ease.

Language (usage and editing services)

Please write your text in good English (American or British usage is accepted, but not a mixture of these). Authors who feel their English language manuscript may require editing to eliminate possible grammatical or spelling errors and to conform to correct scientific English may wish to use the English Language Editing service available from Elsevier's Author Services.

Submission

Our online submission system guides you stepwise through the process of entering your article details and uploading your files. The system converts your article files to a single PDF file used in the peer-review process. Editable files (e.g., Word, LaTeX) are required to typeset your article for final publication. All correspondence, including notification of the Editor's decision and requests for revision, is sent by e-mail.

Submit your article

Please submit your article via https://www.editorialmanager.com/fsigen.

Additional information

Please note that articles that are sent to the author for revision need to be returned within four months. A reminder will be sent in the third month. Any articles that are sent after the fourth month period of revision will be considered a re-submission.

Preparation

Queries

For questions about the editorial process (including the status of manuscripts under review) or for technical support on submissions, please visit our Support Center.

NEW SUBMISSIONS

Submission to this journal proceeds totally online and you will be guided stepwise through the creation and uploading of your files. The system automatically converts your files to a single PDF file, which is used in the peer-review process.
As part of the Your Paper Your Way service, you may choose to submit your manuscript as a single file to be used in the refereeing process. This can be a PDF file or a Word document, in any format or lay-out that can be used by referees to evaluate your manuscript. It should contain high enough quality figures for refereeing. If you prefer to do so, you may still provide all or

some of the source files at the initial submission. Please note that individual figure files larger than 10 MB must be uploaded separately.

References

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the article number or pagination must be present. Use of DOI is highly encouraged. The reference style used by the journal will be applied to the accepted article by Elsevier at the proof stage. Note that missing data will be highlighted at proof stage for the author to correct.

Formatting requirements

There are no strict formatting requirements but all manuscripts must contain the essential elements needed to convey your manuscript, for example Abstract, Keywords, Introduction, Materials and Methods, Results, Conclusions, Artwork and Tables with Captions.
If your article includes any Videos and/or other Supplementary material, this should be included in your initial submission for peer review purposes.
Divide the article into clearly defined sections.

Please ensure your paper has consecutive line numbering - this is an essential peer review requirement.

Figures and tables embedded in text

Please ensure the figures and the tables included in the single file are placed next to the relevant text in the manuscript, rather than at the bottom or the top of the file. The corresponding caption should be placed directly below the figure or table.

Peer review

This journal operates a double anonymized review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper.

The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final. Editors are not involved in decisions about papers which they have written themselves or have been written by family members or colleagues or which relate to products or services in which the editor has an interest. Any such submission is subject to all of the journal's usual procedures, with peer review handled independently of the relevant editor and their research groups. More information on types of peer review.

Double anonymized review

This journal uses double anonymized review, which means the identities of the authors are concealed from the reviewers, and vice versa. More information is available on our website. To facilitate this, please include the following separately:
Title page (with author details): This should include the title, authors' names, affiliations, acknowledgements and any Declaration of Interest statement, and a complete address for the corresponding author including an e-mail address.
Anonymized manuscript (no author details): The main body of the paper (including the references, figures, tables and any acknowledgements) should not include any identifying information, such as the authors' names or affiliations.

REVISED SUBMISSIONS

Use of word processing software

Regardless of the file format of the original submission, at revision you must provide us with an editable file of the entire article. Keep the layout of the text as simple as possible. Most formatting codes will be removed and replaced on processing the article. The electronic text should be prepared in a way very similar to that of conventional manuscripts (see also the Guide to Publishing with Elsevier). See also the section on Electronic artwork.
To avoid unnecessary errors you are strongly advised to use the 'spell-check' and 'grammar-check' functions of your word processor.

Article structure

Manuscripts in general should be organized in the following order:
- Title (should be clear, descriptive and not too long)
- Name(s) of author(s)
- Abstract, which should be clear, descriptive and not longer than 400 words
- Keywords, normally 3-6 items
- Introduction
- Material studied, methods, techniques
- Results
- Discussion
- Conclusion
- Acknowledgments
- References

Essential title page information

• *Title.* Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.
• *Author names and affiliations.* Please clearly indicate the given name(s) and family name(s) of each author and check that all names are accurately spelled. You can add your name between parentheses in your own script behind the English transliteration. Present the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lower-case superscript letter immediately after the author's name and in front of the appropriate address. Provide the full postal address of each affiliation, including the country name and, if available, the e-mail address of each author.
• *Corresponding author.* Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. This responsibility includes answering any future queries about Methodology and Materials. Ensure that the e-mail address is given and that contact details are kept up to date by the corresponding author.
• *Present/permanent address.* If an author has moved since the work described in the article was done, or was visiting at the time, a 'Present address' (or 'Permanent address') may be indicated as a footnote to that author's name. The address at which the author actually did the work must be retained as the main, affiliation address. Superscript Arabic numerals are used for such footnotes.

Highlights

Highlights are mandatory for this journal as they help increase the discoverability of your article via search engines. They consist of a short collection of bullet points that capture the novel results of your research as well as new methods that were used during the study (if any). Please have a look at the examples here: example Highlights.

Highlights should be submitted in a separate editable file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point).

Graphical abstract

Although a graphical abstract is optional, its use is encouraged as it draws more attention to the online article. The graphical abstract should summarize the contents of the article in a concise, pictorial form designed to capture the attention of a wide readership. Graphical abstracts should be submitted as a separate file in the online submission system. Image size: Please provide an image with a minimum of 531 × 1328 pixels (h × w) or proportionally more. The image should be readable at a size of 5 × 13 cm using a regular screen resolution of 96 dpi. Preferred file types: TIFF, EPS, PDF or MS Office files. You can view Example Graphical Abstracts on our information site.
Authors can make use of Elsevier's Illustration Services to ensure the best presentation of their images and in accordance with all technical requirements.

Formatting of funding sources

List funding sources in this standard way to facilitate compliance to funder's requirements:

Funding: This work was supported by the National Institutes of Health [grant numbers xxxx, yyyy]; the Bill & Melinda Gates Foundation, Seattle, WA [grant number zzzz]; and the United States Institutes of Peace [grant number aaaa].

It is not necessary to include detailed descriptions on the program or type of grants and awards. When funding is from a block grant or other resources

available to a university, college, or other research institution, submit the name of the institute or organization that provided the funding.

If no funding has been provided for the research, it is recommended to include the following sentence:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Units

Follow internationally accepted rules and conventions: use the international system of units (SI). If other units are mentioned, please give their equivalent in SI.

Footnotes

Footnotes should be used sparingly. Number them consecutively throughout the article. Many word processors build footnotes into the text, and this feature may be used. Should this not be the case, indicate the position of footnotes in the text and present the footnotes themselves separately at the end of the article.

Artwork

Electronic artwork

General points
• Make sure you use uniform lettering and sizing of your original artwork.
• Preferred fonts: Arial (or Helvetica), Times New Roman (or Times), Symbol, Courier.
• Number the illustrations according to their sequence in the text.
• Use a logical naming convention for your artwork files.
• Indicate per figure if it is a single, 1.5 or 2-column fitting image.
• For Word submissions only, you may still provide figures and their captions, and tables within a single file at the revision stage.
• Please note that individual figure files larger than 10 MB must be provided in separate source files.

A detailed guide on electronic artwork is available.

You are urged to visit this site; some excerpts from the detailed information are given here.

Formats

Regardless of the application used, when your electronic artwork is finalized, please 'save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):

EPS (or PDF): Vector drawings. Embed the font or save the text as 'graphics'.

TIFF (or JPG): Color or grayscale photographs (halftones): always use a minimum of 300 dpi.

TIFF (or JPG): Bitmapped line drawings: use a minimum of 1000 dpi.

TIFF (or JPG): Combinations bitmapped line/half-tone (color or grayscale): a minimum of 500 dpi is required.

Please do not:

• Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); the resolution is too low.

• Supply files that are too low in resolution.

• Submit graphics that are disproportionately large for the content.

Color artwork

Please make sure that artwork files are in an acceptable format (TIFF (or JPEG), EPS (or PDF), or MS Office files) and with the correct resolution. If, together with your accepted article, you submit usable color figures then Elsevier will ensure, at no additional charge, that these figures will appear in color online (e.g., ScienceDirect and other sites) regardless of whether or not these illustrations are reproduced in color in the printed version. For color reproduction in print, you will receive information regarding the costs from Elsevier after receipt of your accepted article. Please indicate your preference for color: in print or online only. Further information on the preparation of electronic artwork.

Figure captions

Ensure that each illustration has a caption. A caption should comprise a brief title (not on the figure itself) and a description of the illustration. Keep

text in the illustrations themselves to a minimum but explain all symbols and abbreviations used.

Tables

Please submit tables as editable text and not as images. Tables can be placed either next to the relevant text in the article, or on separate page(s) at the end. Number tables consecutively in accordance with their appearance in the text and place any table notes below the table body. Be sparing in the use of tables and ensure that the data presented in them do not duplicate results described elsewhere in the article. Please avoid using vertical rules and shading in table cells.

References

Reference links

Increased discoverability of research and high quality peer review are ensured by online links to the sources cited. In order to allow us to create links to abstracting and indexing services, such as Scopus, CrossRef and PubMed, please ensure that data provided in the references are correct. Please note that incorrect surnames, journal/book titles, publication year and pagination may prevent link creation. When copying references, please be careful as they may already contain errors. Use of the DOI is highly encouraged.

A DOI is guaranteed never to change, so you can use it as a permanent link to any electronic article. An example of a citation using DOI for an article not yet in an issue is: VanDecar J.C., Russo R.M., James D.E., Ambeh W.B., Franke M. (2003). Aseismic continuation of the Lesser Antilles slab beneath northeastern Venezuela. Journal of Geophysical Research, https://doi.org/10.1029/2001JB000884. Please note the format of such citations should be in the same style as all other references in the paper.

Data references

This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements:

author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

Reference management software

Most Elsevier journals have their reference template available in many of the most popular reference management software products. These include all products that support Citation Style Language styles, such as Mendeley. Using citation plug-ins from these products, authors only need to select the appropriate journal template when preparing their article, after which citations and bibliographies will be automatically formatted in the journal's style. If no template is yet available for this journal, please follow the format of the sample references and citations as shown in this Guide. If you use reference management software, please ensure that you remove all field codes before submitting the electronic manuscript. More information on how to remove field codes from different reference management software.

Reference formatting

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the article number or pagination must be present. Use of DOI is highly encouraged. The reference style used by the journal will be applied to the accepted article by Elsevier at the proof stage. Note that missing data will be highlighted at proof stage for the author to correct. If you do wish to format the references yourself they should be arranged according to the following examples:

Reference style

Text: Indicate references by number(s) in square brackets in line with the text. The actual authors can be referred to, but the reference number(s) must always be given.

Example: '..... as demonstrated [3,6]. Barnaby and Jones [8] obtained a different result ....'

List: Number the references (numbers in square brackets) in the list in the order in which they appear in the text.

Examples:

Reference to a journal publication:

[1] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, J. Sci. Commun. 163 (2010) 51–59. https://doi.org/10.1016/j.Sc.2010.00372.

Reference to a journal publication with an article number:

[2] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, 2018. The art of writing a scientific article. Heliyon. 19, e00205. https://doi.org/10.1016/j.heliyon.2018.e00205.

Reference to a book:

[3] W. Strunk Jr., E.B. White, The Elements of Style, fourth ed., Longman, New York, 2000.

Reference to a chapter in an edited book:

[4] G.R. Mettam, L.B. Adams, How to prepare an electronic version of your article, in: B.S. Jones, R.Z. Smith (Eds.), Introduction to the Electronic Age, E-Publishing Inc., New York, 2009, pp. 281–304.

Reference to a website:

[5] Cancer Research UK, Cancer statistics reports for the UK. http://www.cancerresearchuk.org/aboutcancer/statistics/cancerstatsreport/, 2003 (accessed 13 March 2003).

Reference to a dataset:

[dataset] [6] M. Oguro, S. Imahiro, S. Saito, T. Nakashizuka, Mortality data for Japanese oak wilt disease and surrounding forest compositions, Mendeley Data, v1, 2015. https://doi.org/10.17632/xwj98nb39r.1.

Reference to software:

[7] E. Coon, M. Berndt, A. Jan, D. Svyatsky, A. Atchley, E. Kikinzon, D. Harp, G. Manzini, E. Shelef, K. Lipnikov, R. Garimella, C. Xu, D. Moulton, S. Karra, S. Painter, E. Jafarov, S. Molins, Advanced Terrestrial Simulator (ATS) v0.88 (Version 0.88), Zenodo, March 25, 2020. https://doi.org/10.5281/zenodo.3727209

Citation in text

Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be

given in full. Unpublished results including publications in preparation, and personal communications should not be included in the reference list, but mentioned in the text as e.g. "J. Smith, unpublished observation" or "J. Smith, personal communication". Citation of a reference as "in press" implies that the item has been accepted for publication. Please be prepared to provide a copy of a reference "in press" upon request to the editor.

Citations to Journal or ISFG Guidelines
Please note: any articles which refer to FSIGEN or ISFG guidelines must include these in the body of the text and not as a citation.

Video

Elsevier accepts video material and animation sequences to support and enhance your scientific research. Authors who have video or animation files that they wish to submit with their article are strongly encouraged to include links to these within the body of the article. This can be done in the same way as a figure or table by referring to the video or animation content and noting in the body text where it should be placed. All submitted files should be properly labeled so that they directly relate to the video file's content. In order to ensure that your video or animation material is directly usable, please provide the file in one of our recommended file formats with a preferred maximum size of 150 MB per file, 1 GB in total. Video and animation files supplied will be published online in the electronic version of your article in Elsevier Web products, including ScienceDirect. Please supply 'stills' with your files: you can choose any frame from the video or animation or make a separate image. These will be used instead of standard icons and will personalize the link to your video data. For more detailed instructions please visit our video instruction pages. Note: since video and animation cannot be embedded in the print version of the journal, please provide text for both the electronic and the print version for the portions of the article that refer to this content.

Data visualization

Include interactive data visualizations in your publication and let your readers interact and engage more closely with your research. Follow the

instructions here to find out about available data visualization options and how to include them with your article.

Supplementary material

Supplementary material such as applications, images and sound clips, can be published with your article to enhance it. Submitted supplementary items are published exactly as they are received (Excel or PowerPoint files will appear as such online). Please submit your material together with the article and supply a concise, descriptive caption for each supplementary file. If you wish to make changes to supplementary material during any stage of the process, please make sure to provide an updated file. Do not annotate any corrections on a previous version. Please switch off the 'Track Changes' option in Microsoft Office files as these will appear in the published version.

Research data

This journal encourages and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles. Research data refers to the results of observations or experimentation that validate research findings. To facilitate reproducibility and data reuse, this journal also encourages you to share your software, code, models, algorithms, protocols, methods and other useful materials related to the project.

Below are a number of ways in which you can associate data with your article or make a statement about the availability of your data when submitting your manuscript. If you are sharing data in one of these ways, you are encouraged to cite the data in your manuscript and reference list. Please refer to the "References" section for more information about data citation. For more information on depositing, sharing and using research data and other relevant research materials, visit the research data page.

Data linking

If you have made your research data available in a data repository, you can link your article directly to the dataset. Elsevier collaborates with a number of repositories to link articles on ScienceDirect with relevant repositories,

giving readers access to underlying data that gives them a better understanding of the research described.

There are different ways to link your datasets to your article. When available, you can directly link your dataset to your article by providing the relevant information in the submission system. For more information, visit the database linking page.

For supported data repositories a repository banner will automatically appear next to your published article on ScienceDirect.

In addition, you can link to relevant data or entities through identifiers within the text of your manuscript, using the following format: Database: xxxx (e.g., TAIR: AT1G01020; CCDC: 734053; PDB: 1XFN).

Mendeley Data

This journal supports Mendeley Data, enabling you to deposit any research data (including raw and processed data, video, code, software, algorithms, protocols, and methods) associated with your manuscript in a free-to-use, open access repository. During the submission process, after uploading your manuscript, you will have the opportunity to upload your relevant datasets directly to *Mendeley Data*. The datasets will be listed and directly accessible to readers next to your published article online.

For more information, visit the Mendeley Data for journals page.

Data statement

To foster transparency, we encourage you to state the availability of your data in your submission. This may be a requirement of your funding body or institution. If your data is unavailable to access or unsuitable to post, you will have the opportunity to indicate why during the submission process, for example by stating that the research data is confidential. The statement will appear with your published article on ScienceDirect. For more information, visit the Data Statement page.