

UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO

MARIANA MEIRA SCUDELER

**CARACTERIZAÇÃO DO GENE *ESR1* EM POPULAÇÕES NATIVAS
DA AMÉRICA DO SUL**

UBERABA

2024

MARIANA MEIRA SCUDELER

**CARACTERIZAÇÃO DO GENE *ESR1* EM POPULAÇÕES NATIVAS
DA AMÉRICA DO SUL**

Dissertação apresentada ao curso de Pós-Graduação em Ciências da Saúde, área de concentração “Medicina Translacional” da Universidade Federal do Triângulo Mineiro, como requisito para obtenção do título de Mestre em Ciências da Saúde

Orientador: Prof^a. Dra. Fernanda Rodrigues Soares

**Uberaba
2024**

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

S442c Scudeler, Mariana Meira
Caracterização do gene *ESR1* em populações nativas da América do Sul
/ Mariana Meira Scudeler. -- 2024.
56 f. : il., fig., graf., tab.

Dissertação (Mestrado em Ciências da Saúde) -- Universidade Federal
do Triângulo Mineiro, Uberaba, MG, 2024
Orientadora: Profa. Dra. Fernanda Rodrigues Soares

1. Neoplasias da mama - Genética. 2. Genes - Efeitos dos fármacos.
3. Farmacogenética. 4. Hereditariedade - Genética. I. Soares, Fernanda
Rodrigues. II. Universidade Federal do Triângulo Mineiro. III. Título.

CDU 618.19-006:575

CARACTERIZAÇÃO DO GENE *ESR1* EM POPULAÇÕES NATIVAS DA AMÉRICA DO SUL

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Saúde, área de concentração “Medicina Translacional” (Linha de Pesquisa: Aspectos Clínicos, Diagnósticos e Terapêuticos das Doenças) da Universidade Federal do Triângulo Mineiro como requisito parcial para obtenção do título de mestre.

Uberaba, 17 de julho de 2024

Banca Examinadora:

Dra. FERNANDA RODRIGUES SOARES – Orientadora
Universidade Federal do Triângulo Mineiro

Dra. FERNANDA SOUZA GOMES KEHDY
IOC / Fiocruz

Dra. MARIANA RODRIGUES BOTTON
Hospital das Clínicas de Porto Alegre



Documento assinado eletronicamente por **FERNANDA RODRIGUES SOARES, Professor do Magistério Superior**, em 19/07/2024, às 11:47, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 165, de 16 de junho de 2023](#).



Documento assinado eletronicamente por **Mariana Rodrigues Botton, Usuário Externo**, em 19/07/2024, às 17:34, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 165, de 16 de junho de 2023](#).



Documento assinado eletronicamente por **Fernanda de Souza Gomes Kehdy, Usuário Externo**, em 13/08/2024, às 16:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 215, de 16 de julho de 2024](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufmt.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1293931** e o código CRC **061CCC01**.

Ao meu pai Oxóssi. Okê aro, meu pai.
À minha mãe Oxum. Ora yeyê ô, mamãe.

AGRADECIMENTOS

A quem é minha direção, meu coração e me dá o chão que piso aqui nessa terra: minha profunda e sincera gratidão a Deus, minha mãe Oxum, meu pai Oxóssi, meu anjo da guarda e meus guias espirituais. Salve os preto-velhos, os caboclos, baianos, erês, marinheiros, boiadeiros, ciganos, exus e pombas-gira que guardam com tanto amor o meu caminhar.

Aos meus pais, Mônia e Renato, por acreditarem tanto em mim e me incentivarem a voar. Tudo que conquistei nessa vida foi por vocês e pelo esforço, sacrifício, amor e encorajamento de vocês. Quanto orgulho eu sinto em carregar pedaços seus em quem eu sou. Mãe, aprendi com você a coragem; aprendi a ser Mulher, nunca desistir e sempre lutar pelo que eu acredito. Pai, você me ensinou o encanto e a sensibilidade; me ensinou a rir e nunca deixar a vida me endurecer. A minha irmã Camila que além de irmã é minha melhor amiga, minha protetora e protegida, meu grande orgulho e companheira de vida. Amo vocês além da vida. Sou grata além da vida por tanta sorte.

Aos meus avós: vó Cida Rocha, vô Ducca, vó Cidinha e vô Carlão (em memória). Vô, perder você com certeza foi a coisa mais difícil e doída que passei nessa caminhada. Sinto saudade todo dia, mas também sinto você comigo todo dia, e sei que você tá orgulhoso dos meus passos. Vocês são os meus grandes, grandes amores, e também com muito orgulho carrego vocês em quem eu sou. Que feliz eu sou pela sorte de uma vida com vocês, por poder cuidar e ser cuidada por vocês. É também pelo seu esforço e amor que estou aqui. Amo vocês infinitamente, de todo coração.

Aos meus padrinhos, Aline e Ronaldo, que sempre estiveram comigo. Também tenho muito orgulho em carregar vocês em quem eu sou. À minha afilhada Giovanna, o amor da minha vida, e meus primos Léo e Lucas que eu tanto amo. Que delícia caminhar a vida com vocês e dividir raízes e histórias com vocês. Não vejo a hora de viver ainda mais histórias juntos.

Às famílias Rocha, Scudeler, Meira e Gonçalves. Tenho muito orgulho, gratidão e amor pelas minhas raízes. Obrigada. Muitas vidas, histórias e caminhos me trouxeram até aqui.

À minha orientadora Fernanda que sempre, sempre foi muito mais que orientadora. Você não tem como mensurar o orgulho, a gratidão e amor que tenho por você e tudo que caminhamos juntas desde 2019. É um privilégio aprender e dividir tanto com você, ciência e vida. Eu sei no meu coração que nosso laço é daqueles que fica, não importa onde estejamos. Obrigada pra sempre, minha orientadora e amiga, por acreditar em mim, por ser tão cuidadosa e me ensinar tanto. Temos muito pela frente.

Ao Caíque, meu grande amor e amigo, por ser tão amoroso, carinhoso e confiante. Dividimos toda essa jornada; sou sempre grata pelo nosso caminho ter se cruzado. Sem você este trabalho não seria possível. Te amo muito. Bruno e Gui, vocês também foram essenciais; sempre disponíveis, carinhosos, amigos. Obrigada de coração. A Maria Laura, que chegou na minha vida antes mesmo da gente imaginar que viveríamos essa jornada juntas; nós nos cuidamos, nos descobrimos e caminhamos muito juntas, minha amiga. Que bom. Te amo muito. A todos os integrantes do Grupo de Pesquisa em Genética de Populações e Farmacogenética: Henrique, Doris, Murilo, Karwhory, Jéssica, Giovanna, Vitória, Tiffany e Arthur, que muito mais que companheiros de trabalho sempre foram meus grandes amigos e minha família científica. Obrigada por tudo.

À Nathalia, Danielle, Helena e Vitória, minha família muito além do sangue. Com vocês eu sonho, me encontro e tenho lugar no mundo. Sou adulta e sou criança. Escolho vocês todos os dias. Amo vocês infinito.

A todos os amigos e companheiros que tive ao longo do meu caminhar. Em especial, obrigada à Mariane e Bruna, que compartilharam tanto do mestrado comigo, e ao Moisés, sempre musicalmente presente ao decorrer de toda essa jornada.

A Uberaba, que me foi casa e solo fértil para ser. Ao Terreiro de Umbanda Sagrada Ogum Beira Mar e Terreiro de Umbanda Vovó Maria Congá, que foram e são o chão sagrado que eu piso.

A todos os professores, técnicos e terceirizados da Universidade Federal do Triângulo Mineiro, por toda a contribuição para a minha formação. Sem vocês nada seria possível.

A mim, que nunca desisti de mim.

RESUMO

Segundo a Organização Mundial de Saúde, o câncer de mama é a primeira ou segunda principal causa de morte entre mulheres em 95% dos países. O estrogênio é considerado um dos fatores de risco associados ao desenvolvimento da doença; é formado principalmente através da ação da enzima aromatase, e sua ação no organismo se dá, em especial, a partir da ligação do hormônio ao receptor de estrogênio ER α , codificado pelo gene *ESR1*. O gene faz parte da via dos anti-estrógenos, que é alvo dos fármacos inibidores da aromatase, cujo mecanismo de ação visa diminuir a concentração de estrogênio no organismo; no entanto, sabe-se que a resposta a tratamentos pode-se alterar a partir de alguns fatores, dentre eles o perfil genético do paciente. A farmacogenética é área que estuda a associação entre os polimorfismos genéticos e a variabilidade interindividual na resposta a fármacos. Além disso, a ancestralidade é também considerada um fator importante neste contexto, isto porque a frequência de polimorfismos nos genes de interesse farmacogenético se alteram ao se comparar diferentes populações. Contudo, populações como as Nativo-americanas, africanas e miscigenadas são sub-representadas nos estudos genéticos; isso é de extrema relevância ao se considerar que essas diferenças podem levar a implicações clínicas distintas para diferentes populações. Com isso, este trabalho tem como objetivo inferir a ancestralidade genômica subcontinental e as frequências alélicas de variantes do gene *ESR1* em 2504 indivíduos de 26 populações ao redor do mundo do Projeto *1000 Genomes* e 150 Nativos do Peru, cujos genomas são dados originais ainda não analisados na literatura em relação ao gene. A partir da inferência da ancestralidade e as frequências, realizar análises de regressão e correlação entre tais dados. Os resultados demonstram a separação subcontinental dos componentes africanos, asiáticos, europeus e nativos (W-AFR, E-AFR, W-EAS, E-EAS, SAS, N-EUR, S-EUR, NAT e NAT2). No total, foram encontradas 145 associações ($p < 0,01$), sendo que 73,1% eram positivas. As frequências de 73 variantes clinicamente relevantes foram associadas aos 9 componentes, com destaque aos componentes africanos para os quais foram observadas mais da metade das associações. Em especial, ressalta-se que foi possível observar associações subcontinentais em relação à maioria das variantes associadas aos componentes nativos, asiáticos e europeus. Além disso, observou-se que cerca de metade das variantes estão em regiões intrônicas, frisando a importância de se estudar genomas completos ao invés de exomas. Nossos resultados apontam para a relevância das diferenças subcontinentais para o design de painéis farmacogenéticos e reforçam os perigos de se extrapolar dados genéticos de uma população para outra. No contexto da prática clínica, os resultados de associação entre variantes clinicamente relevantes do *ESR1* e os componentes de ancestralidade subcontinental podem ser utilizados para guiar as decisões clínicas sobre um paciente na ausência de seus dados genéticos. A inclusão de populações negligenciadas em estudos genômicos e farmacogenômicos é de suma importância para que o acesso aos avanços científicos seja democrático, tão como a implementação Farmacogenética futuramente seja mais igualitária e eficaz, adequada para as especificidades de cada população.

Palavra-chave: Farmacogenética. Câncer de mama. *ESR1*. Ancestralidade.

ABSTRACT

According to the World Health Organization, breast cancer is the first or second leading cause of death among women in 95% of countries. Estrogen is considered to be one of the risk factors associated with the development of the disease; it is formed mainly through the action of the aromatase enzyme, and its action in the organism takes place, in particular, from the hormone's binding to the estrogen receptor ER α , encoded by the *ESR1* gene. The gene is part of the anti-estrogen pathway, which is the target of aromatase inhibitor drugs, whose mechanism of action aims to reduce the concentration of estrogen in the body; however, it is known that the response to treatments may be affected by several factors, including the patient's genetic profile. Pharmacogenetics is the field that investigates the association between genetic polymorphisms and inter-individual variability in the response to drugs. Ancestry is also considered an important factor in this context, as the frequency of polymorphisms in genes of pharmacogenetic interest varies when comparing different populations. Nevertheless, Native American, African, and mixed populations are under-represented in genetic studies; this is extremely relevant when considering that these differences can lead to different clinical implications for different populations. Therefore, this study aims to infer the genomic subcontinental ancestry and allele frequencies of *ESR1* gene variants in 2504 individuals from 26 populations worldwide from the 1000 Genomes Project and 150 Natives from Peru, whose genomes are original data not yet analyzed in the literature regarding this gene. Regression and correlation analyses between these data were carried out based on the inference of ancestry and frequencies. Our findings demonstrate a subcontinental differentiation of African, Asian, European, and Native components (W-AFR, E-AFR, W-EAS, E-EAS, SAS, N-EUR, S-EUR, NAT, and NAT2). A total of 145 associations were found ($p < 0.01$), 73.1% of which were positive. The frequencies of 73 clinically relevant variants were associated with the nine components, particularly the African components, for which more than half of the associations were observed. In particular, subcontinental associations were observed for most of the variants associated with the native, Asian, and European components. Additionally, it was observed that nearly half the variants are in intronic regions, highlighting the importance of studying whole genomes rather than exomes. These results point to the relevance of subcontinental differences for the design of pharmacogenetic panels and reinforce the dangers of extrapolating genetic data from one population to another. In clinical practice, association results between clinically relevant *ESR1* variants and subcontinental ancestry components could be used to guide clinical decisions on a patient in the absence of their own genetic data. Including neglected populations in genomic and pharmacogenomic studies is paramount if access to scientific advances is to be democratic and if pharmacogenetic implementation is to be more egalitarian and effective in the future, tailored to the specificities of each population.

Keywords: Pharmacogenetics. Breast cancer. *ESR1*. Ancestry.

Sumário

1. Introdução	9
2. Revisão da Literatura	12
2.1 Farmacogenética e farmacogenômica	12
2.2 <i>ESR1</i>	14
2.3 Variabilidade interétnica e Ancestralidade Genômica	16
3. Hipótese	21
4. Justificativa	21
5. Objetivos	22
5.1 Objetivo Geral	22
5.2 Objetivos Específicos	22
6. Metodologia	23
7. Resultados e Discussão	28
7.1 Ancestralidade	28
7.2 Anotação das variantes	34
7.3 Associação entre ancestralidade e frequências alélicas e anotação de variantes	36
7.4 Variantes listadas no ClinVar e PharmGKB	38
8. Conclusão	45
Referências	47
Apêndice A	52

1. Introdução

Segundo a Organização Mundial da Saúde, o câncer de mama é a primeira ou segunda principal causa de morte entre mulheres em 95% dos países mundiais. Até 2040, estimam-se 3 milhões de novos casos e 1 milhão de mortes por ano ao redor do mundo (WORLD HEALTH ORGANIZATION, 2023). Múltiplos fatores estão associados ao desenvolvimento da doença: genéticos, como devido a mutações nos genes *BRCA1* e *BRCA2*, que são importantes na codificação de proteínas supressoras de tumor; ambientais/comportamentais, como a partir do consumo de álcool, uma dieta rica em gorduras e sedentarismo; e hormonais, como a exposição ao estrogênio, seja ela endógena, ou seja, quando o hormônio é produzido pelos ovários, ou exógena – a partir do uso de contraceptivos orais e terapias de reposição hormonal (CHAN et al., 2020; SUN et al., 2017). O hormônio atua principalmente através da sua ligação com $ER\alpha$ e $ER\beta$, receptores intracelulares específicos codificados por *ESR1* e *ESR2*, respectivamente. $ER\alpha$ é o principal meio pelo qual a ação oncogênica do estrogênio ocorre, isto devido à ativação de genes promotores de proliferação celular e de diminuição de apoptose promovida pelo receptor (LIANG; SHANG, 2013; SCUDELER et al., 2023). O gene *ESR1*, assim como *HSD17B1*, *CYP19A1* e *ESR2*, fazem parte da via dos inibidores da aromatase (Figura 1), uma classe de fármacos que visa diminuir a produção endógena de estrogênio a partir da inibição da principal enzima catalisadora da biossíntese do hormônio. A terceira geração de tais fármacos é representada pelo letrozol, anastrozol e exemestano. Além disso, a via é também alvo de fármacos moduladores seletivos de receptores do estrogênio (SERMs), como o tamoxifeno, raloxifeno e toremifeno, que atuam como agonistas ou antagonistas dos receptores de estrogênio (LIANG; SHANG, 2013; SUN et al., 2017; WHIRL-CARRILLO et al., 2012).

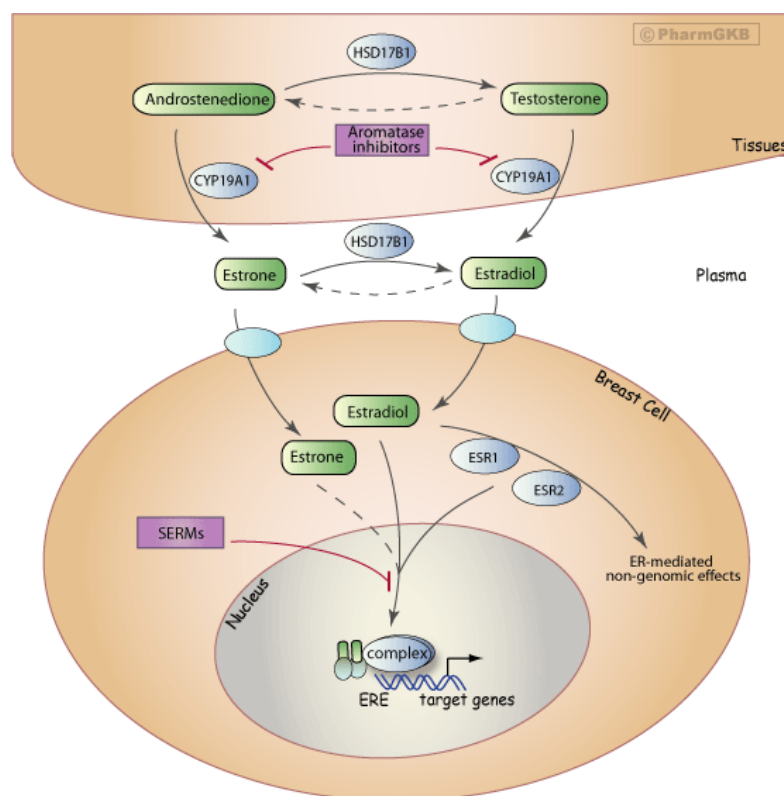


Figura 1: Diagrama representando a via dos inibidores de aromatase em célula mamária. Fonte: (WHIRL-CARRILLO et al., 2012).

A resposta a fármacos, no entanto, não se dá exatamente da mesma forma para todos os pacientes, mesmo que sejam fármacos comprovadamente eficazes; para um pequeno subgrupo destes pacientes podem não funcionar ou causar efeitos adversos, até mesmo a morte. Isto porque existem diversos fatores específicos que podem levar à variabilidade interindividual de resposta, sendo eles ambientais e fisiológicos – geralmente transitórios, ou genéticos, por sua vez permanentes (MA; LU, 2011). Polimorfismos genéticos, que são variações na sequência de DNA presentes com uma frequência maior ou igual a 1%, quando levam a alterações em enzimas, receptores e transportadores envolvidos na farmacocinética e farmacodinâmica de fármacos são considerados possivelmente os principais causadores de variabilidade interindividual (BORGES-OSÓRIO; ROBINSON, 2013; MA; LU, 2011). A farmacogenética (FGx) é a área que surge para buscar entender a associação entre tais polimorfismos presentes nos farmacogenes, que são os genes de interesse farmacogenético, e a variação de resposta à tratamentos farmacológicos (SCUDELER; RODRIGUES-SOARES, 2020). O polimorfismo rs9322336 do gene *ESR1*, por exemplo, foi associado a síndrome musculoesquelética e descontinuação do tratamento em pacientes tratados com

exemestano (HENRY et al., 2013); as reações adversas muscoesqueléticas foram também associadas a rs2234693 e rs9340799 em resposta ao uso de letrozol e anastrozol em um grupo de pacientes leste-asiáticos (WANG et al., 2013). Relata-se, além disso, o desenvolvimento de artralgia associada às 3 variantes citadas a partir da prescrição de letrozol e anastrozol (BORRIE et al., 2020). A etnia é outro fator influenciador no que diz respeito à variabilidade interindividual na resposta a fármacos, visto que a frequência de polimorfismos nos farmacogenes é heterogênea ao se comparar diferentes populações, podendo levar a implicações clínicas específicas (RAMOS et al., 2014; SCUDELER; RODRIGUES-SOARES, 2020; SOSA-MACÍAS et al., 2016; TAN-KOI et al., 2019). No entanto, é importante ressaltar que não necessariamente a etnia autodeclarada de um indivíduo corresponderá a sua ancestralidade genômica – ou seja, suas origens genéticas populacionais (PADHUKASAHASRAM, 2014; SUAREZ-KURTZ et al., 2007). Isto especialmente tratando-se de populações miscigenadas como a brasileira (SUAREZ-KURTZ et al., 2007), que semelhante às da América Latina em geral, possui raízes étnicas oriundas de 3 populações parentais: Africanos, Europeus e Nativo-americanos (SCUDELER; RODRIGUES-SOARES, 2020; SUAREZ-KURTZ; PARRA, 2018). No que diz respeito às implicações clínicas específicas, tem-se a varfarina, um anticoagulante amplamente utilizado na prática clínica, como um dos principais exemplos: sabe-se que a dose necessária para efeito terapêutico é diferente para afro-americanos, brancos e asiáticos, isto devido a diferenças de frequência de variantes dos principais genes envolvidos no seu metabolismo e transporte entre estas populações – *CYP2C9*, *CYP2C19* e *VKORC1* (SCUDELER; RODRIGUES-SOARES, 2020; URBAN, 2010). Recentemente, Moreira et al. (2022) discutiram que um ciclo extra de quimioterapia pode ser necessário para pacientes de ancestralidade nativo-americana com leucemia linfoblástica aguda para se obter a eficácia esperada (MOREIRA et al., 2022). Apesar disso, sabe-se que populações não-europeias, como as Nativo-americanas, são negligenciadas na representação em estudos genéticos e genômicos; o que se observa, na maioria das vezes, é uma super-representação europeia (MCGUIRE et al., 2020; MOREIRA et al., 2022; SCUDELER; RODRIGUES-SOARES, 2020; WOJCIK et al., 2019). A inclusão de populações sub-representadas nos estudos mostra-se bastante importante para o design de painéis de variantes, visto que a seleção de uma variante extremamente rara em determinada população seria uma alocação ineficaz e injustificada de recursos. Com isso, dadas as diferenças interpopulacionais na frequência de polimorfismos em farmacogenes e o

potencial da genotipagem do *ESR1* no que diz respeito à eficácia e segurança do tratamento hormonal para pacientes do câncer de mama, tem-se como objetivo correlacionar a frequência de variantes do gene *ESR1* e a ancestralidade subcontinental de populações mundiais, incluindo dados originais de populações indígenas ainda não avaliadas na literatura neste contexto. Desse modo é possível que futuramente a implementação clínica da farmacogenética seja mais eficaz, assertiva e verdadeiramente democrática para diferentes populações ao redor do mundo.

2. Revisão da Literatura

2.1 Farmacogenética e farmacogenômica

Alguns autores remontam o surgimento da Farmacogenética no ano de 510 a.C., quando Pitágoras descreveu que a ingestão de feijão-fava poderia levar a sérias reações em alguns indivíduos – mas não todos. Durante a Segunda Guerra Mundial, soldados afro-americanos desenvolveram anemia hemolítica aguda secundária ao uso de primaquina, um fármaco antimalárico. Na década de 50 do século XX, pôde-se estabelecer que tais reações, tanto a observada por Pitágoras quanto a hemólise induzida por primaquina dos soldados, estão relacionadas a deficiência de glicose-6-fosfato desidrogenase (G6PD), uma condição causada por mutações no gene *G6PD* do cromossomo X (BELFIELD; TICHY, 2018; BROCKMÖLLER; TZVETKOV, 2008; ENGEN et al., 2006; HAMAD; FELDMAN, 2010; OZDEMIR; SHEAR; KALOW, 2001). A área inicia, então, seu crescimento a partir de 1950 com a constatação de componentes genéticos envolvidos no metabolismo de fármacos e, portanto, na variação de resposta aos tratamentos (HAMAD; FELDMAN, 2010; MA; LU, 2011; URBAN, 2010). Ao se falar das descobertas desta época, o relaxante muscular succinilcolina e o antibiótico antituberculose isoniazida são dois grandes exemplos devido à sua ligação com as enzimas N-acetiltransferases (HAMAD; FELDMAN, 2010; URBAN, 2010). Desde então a FGx se tornou promessa cada vez mais importante dentro da medicina de precisão, visto que os componentes genéticos são parte importante da variabilidade interindividual na resposta aos fármacos (MA; LU, 2011; ZHANG; FINKELSTEIN, 2019). Isto especialmente com os progressos das tecnologias de sequenciamento de nova geração e iniciativas como o Projeto Genoma Humano e o Projeto *1000 Genomes*, a partir dos quais foi possível aprofundar exponencialmente os conhecimentos existentes sobre variação genética e a interação entre genoma, doença e a metabolização de fármacos (CAULFIELD et al., 2009; MA; LU, 2011; SUAREZ-KURTZ; PARRA, 2018). Com

isso nasce um novo termo, a farmacogenômica, que alguns autores diferenciam da farmacogenética ao defini-la como o estudo da variabilidade interindividual a partir do genoma e não de um gene ou genótipo candidato (HAMAD; FELDMAN, 2010; KOZYRA; INGELMAN-SUNDBERG; LAUSCHKE, 2017; MA; LU, 2011; OZDEMIR; SHEAR; KALOW, 2001). Outros autores utilizam os termos como sinônimos (DALY, 2017; LAM, 2019), e assim será feito neste trabalho.

O genoma humano é composto por cerca de 25.000 genes (HAMAD; FELDMAN, 2010) e os polimorfismos de nucleotídeo único (SNPs) são as variantes mais comuns, com mais de 20 milhões descritas. Além dos SNPs outros tipos de variantes também podem afetar o metabolismo de fármacos, como inserções-deleções (INDELs), repetições em tandem, variação no número de cópias (CNVs), dentre outros (LAM, 2019; MA; LU, 2011). As variações genéticas comumente ocorrem em genes codificadores de proteínas-alvo (como receptores e enzimas) ou proteínas envolvidas no transporte e metabolismo de fármacos (i.e. transportadores e enzimas), afetando, portanto, a farmacodinâmica e farmacocinética e, deste modo, alterando a eficácia e segurança do fármaco (LAM, 2019; MA; LU, 2011). As enzimas da superfamília citocromo P450 (CYPs) são codificadas por genes homônimos, e estes são os mais estudados no contexto farmacogenético visto que participam do metabolismo de grande parte dos fármacos e xenobióticos, além de serem altamente polimórficos, apresentando SNPs responsáveis pela maior parte da variabilidade nas atividades enzimáticas. Além disso, também estão descritas deleções, variação no número de cópias e outras alterações (KOZYRA; INGELMAN-SUNDBERG; LAUSCHKE, 2017; LAM, 2019; SCUDELER; RODRIGUES-SOARES, 2020). No entanto, além das alterações de farmacocinética que por muito tempo foram o maior foco dos estudos, variantes que afetam a farmacodinâmica são também bastante importantes, o que tem sido cada vez mais demonstrado (LAM, 2019; MA; LU, 2011; URBAN, 2010). Atualmente na literatura são diversos os achados que associam variantes do gene *ESR1* com alterações na resposta a tratamentos. Segundo Oesterreich e colaboradores, os genótipos CT + TT do rs4870061 foram associados à uma diminuição da densidade mineral óssea em pacientes com câncer de mama tratados com exemestano ou letrozol, assim como o genótipo TT do rs9322335 (OESTERREICH et al., 2015). O alelo C do rs2234693, por sua vez, foi associado a um aumento do risco de dor musculoesquelética em pacientes tratadas com anastrozol e letrozol (WANG et al., 2013). Além disso, os genótipos AG+GG do rs9340799 foi associado a trombose venosa profunda e embolia pulmonar devido ao uso de tamoxifeno (ONITILO et al., 2009). O

objetivo da farmacogenética é justamente evitar reações como estas ao guiar a prescrição farmacológica preventivamente em acordo com o perfil genético do paciente (SCUDELER; RODRIGUES-SOARES, 2020; URBAN, 2010).

Visando abrir caminho para a implementação da FGx e atuando como ferramentas para que os objetivos da área sejam alcançados, algumas iniciativas foram criadas ao decorrer dos anos. O PharmGKB (<https://www.pharmgkb.org/>) é um banco de dados que realiza a análise, curadoria e disseminação das evidências farmacogenéticas disponíveis na literatura (WHIRL-CARRILLO et al., 2012). Além disso, o *Clinical Pharmacogenetics Implementation Consortium* (CPIC) é um consórcio internacional que atua especificamente na criação de guidelines revisadas por pares acerca de variantes genéticas, suas consequências na metabolização de fármacos e a tomada de decisão na prática clínica baseada em evidências (<https://cpicpgx.org>). Os dados que se referem especificamente ao *ESR1*, objeto de estudo deste trabalho, serão discutidos e aprofundados posteriormente no capítulo subsequente.

2.2 *ESR1*

O gene Receptor de Estrogênio 1 – *ESR1* –, também conhecido como *ER*, *ESR*, *Era*, *ESRA*, *ESTRR* e *NR3A1*, está localizado no cromossomo 6 (GRINSHPUN et al., 2023; NATIONAL LIBRARY OF MEDICINE, 2023). Pertence a uma superfamília de receptores hormonais nucleares e, portanto, como seu próprio nome diz, é responsável por codificar o receptor de estrogênio alfa ($ER\alpha$), um fator de transcrição ativado pelo ligante formado por uma estrutura de 595 aminoácidos (GRINSHPUN et al., 2023; GRINSHPUN; SANDUSKY; JESELSON, 2023; LIANG; SHANG, 2013). O receptor é amplamente expresso em múltiplos tecidos como a glândula mamária, os ovários, o útero e, além disso, também no cérebro e nos pulmões (LIANG; SHANG, 2013). Em acordo com Grinshpun e colaboradores (2023), possui um papel importante em mecanismos de progressão tumoral e tumorigênese, especialmente no que diz respeito ao câncer de mama (GRINSHPUN et al., 2023; GRINSHPUN; SANDUSKY; JESELSON, 2023).

Dentre os critérios de classificação do câncer de mama citam-se os subtipos moleculares, em que se observa a expressão (proteica ou gênica, o padrão ouro) de receptores de estrogênio (ER), progesterona (PR) e HER2. A partir desta análise, classifica-se o caso dentre os grupos: Luminal A e B, que possuem expressão de ER e/ou

PR, HER2+, que como denuncia o nome é caracterizado por uma alta expressão de HER2 e, por fim, o triplo-negativo, que por sua vez não expressa nenhum dos receptores (SCUDELER et al., 2023; SUN et al., 2017). Sabe-se que os tumores ER+ são os mais prevalentes, representando cerca de 70-80% dos casos, independente do estágio da doença (GRINSHUPUN et al., 2023; LIANG; SHANG, 2013; SUN et al., 2017; WILL et al., 2023). Com isso e dado seu importante papel na patogênese do câncer de mama, são considerados fatores informativos de prognóstico e biomarcadores preditivos de resposta à terapia endócrina da qual é alvo, sendo este o tratamento padrão empregado para estes casos (SCUDELER et al., 2023). Utiliza-se, por exemplo, moduladores seletivos de ER (SERMs), como o tamoxifeno, e inibidores da aromatase (GRINSHUPUN et al., 2023; SCUDELER et al., 2023).

Segundo o PharmGKB, são duas as vias em que o gene participa. Na Via da Lamotrigina descreve-se que a ativação do ER α pela alta de 17 β -estradiol na gravidez leva à regulação positiva da expressão do *UGT1A4* e aumento do *clearance* do fármaco (MITRA-GHOSH et al., 2020). A segunda – e a que nos é de maior interesse – é a Via dos Inibidores da Aromatase. O estrogênio é biossintetizado a partir dos andrógenos com a ação da enzima aromatase; a partir da sua ligação com os receptores de estrogênio (ER α e ER β), recruta proteínas coativadoras ou correpessoras e juntos se dimerizam, mudam sua conformação e se ligam nos Elementos de Resposta a Estrogênio (EREs), levando a um aumento da expressão de genes associados a proliferação celular do tecido mamário. Moduladores seletivos de receptores de estrogênio atuam regulando este processo de modo a diminuir os efeitos do estrogênio nos processos de proliferação celular (WHIRL-CARRILLO et al., 2012). Na aba “*Variant Annotations*” estão reportadas associações entre uma variante genética e um fenótipo relacionado a fármacos (toxicidade, eficácia, dosagem ou outros). Até então existem descritas 30 anotações, dentre as quais estão as referentes a polimorfismos do *ESR1* e os fármacos inibidores da aromatase e SERMs em relação ao câncer de mama. O genótipo TT do rs9322335 foi associado a diminuição de densidade óssea de mulheres com câncer de mama fazendo uso de letrozol e de exemestano ou letrozol. O mesmo foi observado neste mesmo artigo para o genótipo TT do rs4870061 em relação ao uso de letrozol e CT + TT para o uso de exemestano ou letrozol, além do genótipo AA do rs2813543 em relação ao exemestano (OESTERREICH et al., 2015). Por sua vez, rs9322336 (CC) foi associado ao aumento das chances de ocorrência de toxicidade muscoesquelética em resposta ao exemestano (HENRY et al., 2013). Wang e colaboradores (2013) observaram que os alelos A do rs9340799 e C do

rs2234693 estão associados a um risco aumentado de dor muscoesquelética devido à exposição a anastrozol ou letrozol (WANG et al., 2013). Além disso, os genótipos AG+GG do rs9340799 foram associados ao risco de embolia e trombose em pacientes fazendo uso de tamoxifeno; outro estudo associou o genótipo GG a alterações lipídicas em resposta ao mesmo fármaco (NTUKIDEM et al., 2008; ONITILO et al., 2009). Em “*Clinical Annotations*” estão listadas 12 anotações, que consistem em pares variante-fármaco, sendo possível consultar uma tabela que relaciona os genótipos e seus respectivos impactos fenotípicos em acordo com níveis de evidência. Considerando os pares entre variantes do *ESR1* e inibidores da aromatase ou SERMs no contexto do câncer de mama, até então estão descritas 6 anotações, todos relacionados à categoria fenotípica de toxicidade, com nível 3 de evidência (WHIRL-CARRILLO et al., 2021). Em uma recente revisão sistemática abordando a farmacogenética do câncer de mama, observou-se que duas variantes do *ESR1* (XbaI e PvuII) fazem parte das principais variantes para as quais foram encontradas associações com variação na resposta a tratamento; no entanto, a maioria dos indivíduos estudados em relação ao gene não teve sua etnia informada. Considerando os que tiveram, destacam-se os brancos e leste-asiáticos. Não foram estudadas pacientes Nativo-americanas, médio-orientais, sul-asiáticas e africanas, e dentre as hispânicas somente 74 indivíduos foram estudados (SCUDELER et al., 2023). Nota-se que variantes do gene estão bem estabelecidas na literatura como fatores relevantes para o acesso a tratamentos de maior eficácia e segurança, mas junto a isso ressalta-se uma lacuna étnica e populacional a ser preenchida.

2.3 Variabilidade interétnica e Ancestralidade Genômica

A ideia de que etnia e ancestralidade são fatores relevantes no contexto farmacogenético já é levantada há algum tempo. Segundo Urban (2010), já desde 1982 evidências indicam diferenças interétnicas na resposta a anti-hipertensivos, com indivíduos considerados brancos exibindo uma melhor resposta a beta-bloqueadores em comparação a pretos (URBAN, 2010). Com o desenvolvimento científico no decorrer dos anos, foram descobertas variantes em genes envolvidos no metabolismo de tais fármacos associados à alteração na resposta; observou-se também que indivíduos de ancestralidade europeia eram mais propensos a ser carreadores de tais variantes em comparação a indivíduos de ancestralidade africana, o que pode explicar tal variação, tão como ou juntamente a fatores ambientais. Outro exemplo muito conhecido é o da varfarina, já

aqui mencionado dada às diferentes doses necessárias para a eficácia esperada a depender da etnia do paciente devido a diferenças de frequência nos genes envolvidos no metabolismo do fármaco (SCUDELER; RODRIGUES-SOARES, 2020; URBAN, 2010). Atualmente já se tem muito bem estabelecido que a frequência de variantes genéticas difere em diferentes populações (SCUDELER et al., 2023; SUAREZ-KURTZ; PARRA, 2018; TAN-KOI et al., 2019); processos de seleção natural, deriva genética e pressão ambiental são responsáveis pelo estabelecimento dessas diferenças (PENA et al., 2011; SIRUGO; WILLIAMS; TISHKOFF, 2019; SUAREZ-KURTZ et al., 2007; URBAN, 2010). Mais especificamente no contexto farmacogenético, Ramos e colaboradores (2014) estudaram 1478 indivíduos de 19 populações ao redor do mundo em relação a 1156 variantes genéticas (dentre elas 42 acionáveis clinicamente em acordo com o PharmGKB e FDA) de 212 genes envolvidos na absorção, distribuição, metabolismo e excreção de fármacos (i.e. farmacocinética) (RAMOS et al., 2014; SUAREZ-KURTZ, 2018). Os autores realizaram análises referentes à frequência do alelo menor (MAF), diversidade genética (Fst) e estrutura populacional e a partir de todas as análises pôde-se observar clara diferenciação populacional. Especificamente em relação às variantes clinicamente acionáveis, observou-se ampla variação nas frequências alélicas: rs9923231 do gene *VKORC1* (SAYERS et al., 2022), por exemplo, que é associado à necessidade de alterações na dosagem de varfarina, apresentou frequência que variou de 0,02 a 0,95; outros exemplos associados a antineoplásicos, antibióticos, anticoagulantes e outros estão disponíveis na Figura 2 (RAMOS et al., 2014).

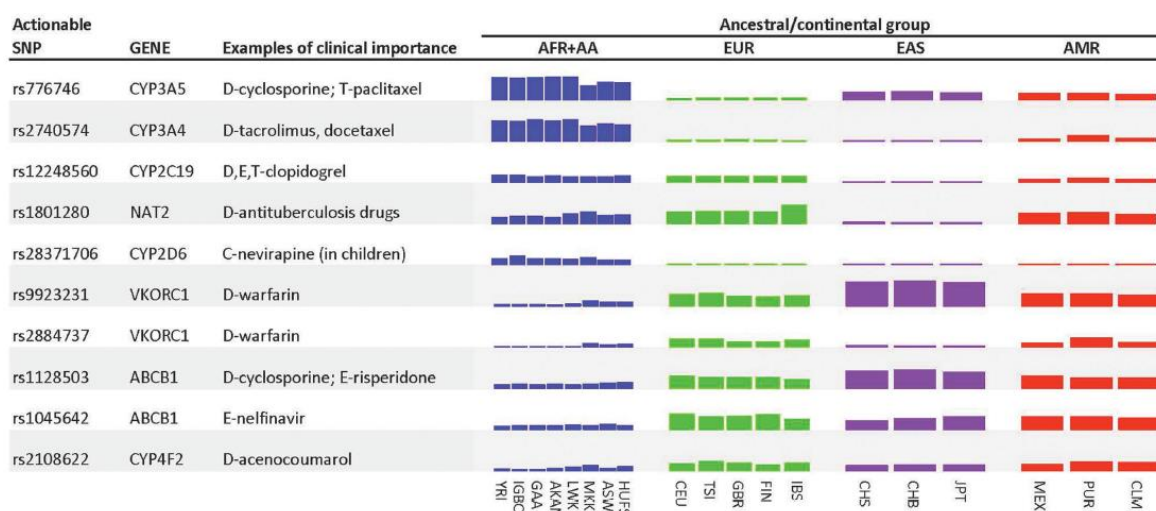


Figura 2: Variantes genéticas clinicamente acionáveis e suas frequências em diferentes populações. Fonte: (RAMOS et al., 2014).

Além disso, um estudo recente analisou 60.706 exomas humanos (Exome Aggregation Consortium (ExAC)) a fim de avaliar a prevalência de variantes funcionais em 806 genes associados a fármacos, o risco interpopulacional de que tais variantes afetem o metabolismo de 1236 fármacos aprovados pela FDA e os 100 fármacos mais prescritos nos Estados Unidos e a probabilidade de um indivíduo ser carreador de uma variante funcional (SCHÄRFE et al., 2017). Os achados descritos pelos autores apontam que 80% dos indivíduos têm chance de ter ao menos uma variante funcional que afete fármacos comumente prescritos e que a frequência de tais variantes varia entre diferentes ancestralidades. Os resultados estão em consonância com os encontrados por Kozyra e colaboradores (2016). Os autores analisaram 1.092 genomas (Projeto 1000 Genomes) e 6.503 exomas (Exome Sequencing Project (ESP)) para avaliar a variabilidade genética de 146 genes envolvidos na farmacocinética de fármacos em populações europeias e africanas. O impacto funcional das variantes foi previsto por análises *in silico*; dentre as 12.152 variantes exônicas encontradas, 6.750 foram consideradas funcionais. No contexto populacional os autores observaram que 72% das variantes nos farmacogenes eram específicas a uma população – 5.100 a europeus e 3.977 a africanos. Em conjunto, os dados mostram considerável variação interpopulacional da variabilidade genética dos farmacogenes estudados (KOZYRA; INGELMAN-SUNDBERG; LAUSCHKE, 2017). Adicionalmente, ressalta-se a importância de se estudar genomas completos neste contexto, visto que a análise de exomas não abrange potenciais variantes regulatórias localizadas em regiões intrônicas (SUAREZ-KURTZ; PARRA, 2018). Dada a variação de frequências de variantes fica explícita a necessidade de se incluir a maior diversidade possível de populações nos estudos genéticos e farmacogenéticos, evitando o extrapolação de dados entre populações, visto que isso pode levar a resultados errôneos (BERGSTRÖM et al., 2020; BUSTAMANTE; DE LA VEGA; BURCHARD, 2011; ENGEN et al., 2006; KOZYRA; INGELMAN-SUNDBERG; LAUSCHKE, 2017; SCHÄRFE et al., 2017; SIRUGO; WILLIAMS; TISHKOFF, 2019; SUAREZ-KURTZ; PARRA, 2018; WOJCIK et al., 2019). Tal necessidade é especialmente importante quando se trata de populações miscigenadas como as da América Latina, que são originárias de três raízes ancestrais comuns: Ameríndios, Africanos e Europeus (HOMBURGER et al., 2015; MARIANA R. BOTTON et al., 2019; SUAREZ-KURTZ et al., 2007).

Sabe-se que diversos estudos utilizam informações de etnia autodeclarada como proxy para ancestralidade (SUAREZ-KURTZ et al., 2007); no entanto, é de extrema

importância que se saiba que não são conceitos equivalentes (ENGEN et al., 2006; PADHUKASAHASRAM, 2014). O conceito de etnia é complexo e dinâmico, abrangendo concepções biológicas, sociais, culturais, históricas, de linguagem e nacionalidade (ENGEN et al., 2006; TAN-KOI et al., 2019). Ancestralidade, por sua vez, consiste na inferência das proporções de ancestralidade oriundas de diferentes populações em relação a todo o genoma a partir de marcadores genéticos (ENGEN et al., 2006; PADHUKASAHASRAM, 2014). Desse modo, não é o ideal extrapolar dados de etnia autodeclarada para definir ancestralidade (TAN-KOI et al., 2019), que atualmente pode ser estimada a partir dados genéticos analisados por softwares como *Structure* e *Admixture*, que utilizam modelos estatísticos, ou a partir de técnicas como análises de clusters ou de componentes principais (PCA) (PADHUKASAHASRAM, 2014). Estudos genéticos comumente consideram populações miscigenadas como originárias de populações parentais de diferentes continentes; no entanto, é possível que a ancestralidade continental se diferencie em subcontinental, em que há especificidade a sub-regiões do continente (GOUVEIA et al., 2020, 2023; HOMBURGER et al., 2015).

Também é um fato muito bem estabelecido na literatura que os estudos genômicos e farmacogenéticos são principalmente realizados com populações europeias e eurodescendentes (BUSTAMANTE; DE LA VEGA; BURCHARD, 2011; MOREIRA et al., 2022; SCUDELER; RODRIGUES-SOARES, 2020; SIRUGO; WILLIAMS; TISHKOFF, 2019; SUAREZ-KURTZ; PARRA, 2018; URBAN, 2010; WOJCIK et al., 2019). O *GWAS Catalog* é uma iniciativa que surgiu em 2008 para facilitar o acesso aos estudos e novos dados gerados a partir do rápido crescimento da genômica (SOLLIS et al., 2023); dentre indivíduos estudados nos *genome-wide association studies* (GWAS) listados na plataforma até janeiro de 2019, 78,39% são europeus. Africanos representam 2%, latino-americanos 1,13% e o grupo que soma médio-orientais, ameríndios e nativos da Oceania não chega a atingir 1% (Figura 3; (SIRUGO; WILLIAMS; TISHKOFF, 2019). Populações miscigenadas, africanas, sul-asiáticas e nativas são comumente negligenciadas, inclusive no contexto farmacogenético de estudos de câncer de mama avaliando o gene *ESR1*, de modo que os avanços tidos com o desenvolvimento científico recente não beneficiam a todos igualmente (BUSTAMANTE; DE LA VEGA; BURCHARD, 2011; ENGEN et al., 2006; MOREIRA et al., 2022; SCUDELER et al., 2023; SIRUGO; WILLIAMS; TISHKOFF, 2019; WOJCIK et al., 2019). Através de tais avanços, o papel da genética na eficácia e segurança dos mais diversos tratamentos foi progressivamente estabelecido, mas assim como se evidenciou a sua importância,

evidenciou-se e exacerbou-se a desigualdade no acesso à saúde (SIRUGO; WILLIAMS; TISHKOFF, 2019; URBAN, 2010; WOJCIK et al., 2019).

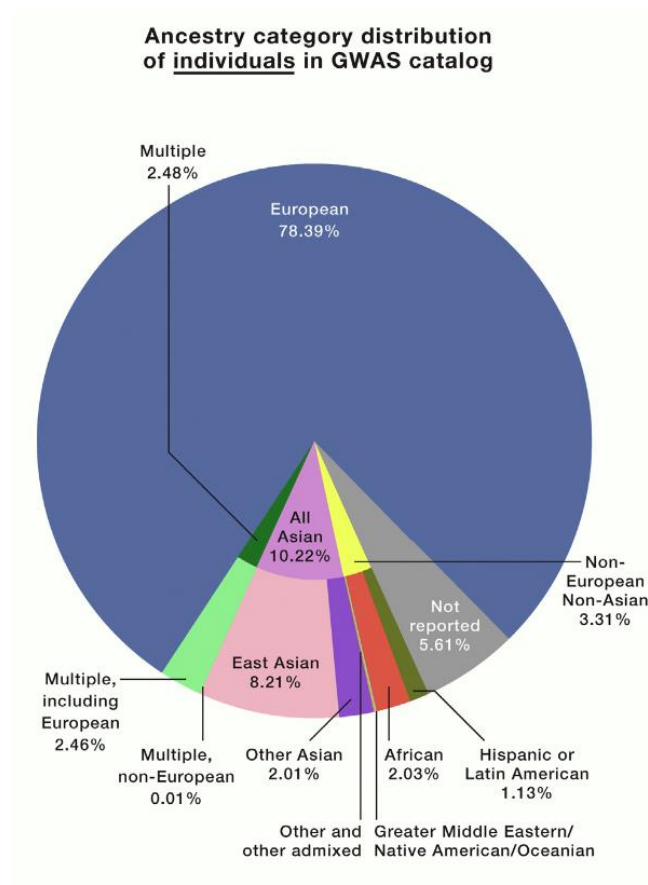


Figura 3: Ancestralidade dos indivíduos estudados nos GWAS listados do *GWAS Catalog* até o ano de 2018. Fonte: Adaptado de (SIRUGO; WILLIAMS; TISHKOFF, 2019).

Iniciativas como a Rede Iberoamericana de Farmacogenética e Farmacogenômica (RIBEF) surgiram para tentar mudar este cenário. A RIBEF foi criada em 2006 pelo Dr. Adrián Llerena com o objetivo de incluir populações latinas nos estudos farmacogenéticos com a participação de dezenas de grupos de pesquisa em diferentes países Ibero-Americanos (RODEIRO et al., 2012; SCUDELER; RODRIGUES-SOARES, 2020; SOSA-MACÍAS et al., 2016). Além disso, foi criada no ano de 2004 no Brasil a Rede Nacional de Farmacogenética (REFARGEN), liderada pelo Dr. Guilherme Suarez-Kurtz. A rede é composta por diversos grupos de pesquisa de diferentes regiões do país buscando manter arquivos de dados farmacogenéticos na população brasileira, guiar a implementação clínica da Farmacogenética assim como promover iniciativas educacionais acerca da área (<https://www.refargen.org.br>) (SCUDELER; RODRIGUES-SOARES, 2020). O crescimento da FGx trouxe uma grande expectativa no que diz

respeito à implementação da medicina personalizada (ENGEN et al., 2006), mas para que isso ocorra de maneira democrática, igualitária e de fato efetiva, é absolutamente essencial que se inclua populações sub-representadas e negligenciadas nos estudos.

3. Hipótese

Considerando o fato de que a frequência de variantes farmacogenéticas se diferencia em diferentes populações, através dos dados do banco *1000 Genomes* e dados originais de Nativos peruanos, hipotetiza-se que serão observadas associações entre a frequência de alelos do gene *ESR1* e ancestralidade subcontinental, em especial a população nativo-americana, ainda subrepresentada nos estudos genéticos e por isso não igualitariamente beneficiada pelos avanços científicos atuais.

4. Justificativa

Até o presente momento, considerando o que temos conhecimento, a sequência do gene *ESR1*, tido como relevante na resposta ao tratamento para câncer de mama com inibidores da aromatase, ainda não foi analisada em populações nativas. Além disso, estudos de associação entre ancestralidade genômica e alelos de interesse farmacogenético são de grande importância para guiar o desenvolvimento de painéis farmacogenéticos assertivos e funcionais para cada população, englobando suas particularidades, e assim permitindo resultados mais confiáveis e certos para uma aplicação farmacogenética democrática na prática clínica.

5. Objetivos

5.1 Objetivo Geral

Caracterizar o gene *ESR1* em populações nativas e miscigenadas da América do Sul e testar a associação entre ancestralidade genômica subcontinental com e frequências alélicas do *ESR1*.

5.2 Objetivos Específicos

- Baixar dados do Projeto *1000 Genomes* referente a variantes do gene *ESR1*;
- Realizar o controle de qualidade bioinformático da sequência do gene *ESR1* em 150 indivíduos Nativo-americanos do Peru;
- Anotar as variantes encontradas em acordo com o dbSNP;
- Inferir a ancestralidade genômica de 2504 indivíduos das 26 populações do Projeto *1000 Genomes*;
- Calcular a frequência dos alelos do gene *ESR1* nas 26 populações do Projeto *1000 Genomes* e nas populações nativas e miscigenadas do Peru;
- Efetuar anotação das variantes das populações de estudo em relação a bancos de dados públicos;
- Testar a associação entre os componentes de ancestralidade subcontinental e a frequência de variantes do gene *ESR1*.

6. Metodologia

O banco de dados público *1000 Genomes* (AUTON et al., 2015) foi fonte dos genomas de 2504 indivíduos de 26 diferentes populações ao redor do mundo. A partir destes indivíduos, foram filtradas as sequências referentes ao gene *ESR1* e genótipos de 1.363.818 SNPs para a realização da análise de ancestralidade *genome-wide*. Além disso, foram incluídos dados originais de sequenciamento do genoma completo de 150 indivíduos Nativo-americanos do Peru que são oriundos das populações Matzes, Moches, Trujillo, Iquitos, Cusco, Chopccas e Uros (Tabela 1).

Tabela 1: Número de indivíduos por população das 33 populações incluídas no estudo.

População	Abreviação	N
Britânicos vivendo na Inglaterra e Escócia	GBR	91
Finlandeses	FIN	99
Descendentes de europeus residentes em Utah	CEU	99
Indivíduos da Toscana, Itália	TSI	107
Ibéricos da Espanha	IBS	107
Japoneses em Tóquio	JPT	104
Chineses Han da região sul	CHS	105
Chineses Han de Pequim	CHB	103
Chineses Dai em Xishuangbanna, China	CDX	93
Vietnamitas Kinh	KHV	99
Indianos Gujarati em Houston, Texas	GIH	103
Indianos Telugu vivendo na Inglaterra	ITU	102
Indivíduos do Tamil do Sri Lanka vivendo na Inglaterra	STU	102
Indivíduos de Bengali, Bangladesh	BEB	86
Punjabi em Lahore, Paquistão	PJL	96
Mandinka em Gâmbia	GWD	113
Mante de Saerra	MSL	85
Yorubas da Nigéria	YRI	108
Esan da Nigéria	ESN	99

Luhya no Quênia	LWK	99
Afro-caribenhos em Barbados	ACB	96
Afro-americanos do sudoeste dos EUA	ASW	61
Peruanos de Lima, Peru	PEL	85
Descendentes de mexicanos em LA	MXL	64
Colombianos de Medellín	CLM	94
Porto-riquenhos	PUR	104
Matzes	MATZES	12
Moches	MOCHES	30
Trujillo	TRUJILLO	16
Iquitos	IQUITOS	16
Cusco	CUSCO	16
Chopccas	CHOPCCAS	30
Uros	UROS	30

Fonte: (AUTON et al., 2015; HARRIS et al., 2018).

A ancestralidade foi inferida pelo software *Admixture* (ALEXANDER; NOVEMBRE; LANGE, 2009). Para a escolha das variantes foram aplicados filtros utilizando o software *Plink* (PURCELL et al., 2007) para cada população separadamente para exclusão de variantes com MAF $< 0,05$ e valor de desequilíbrio de ligação (LD) $r^2 \geq 0,8$. Com isso, as populações foram agrupadas e foram aplicados filtros para a exclusão de duplicatas, variantes que não fossem bialélicas e que possuíam dados faltantes, resultando na seleção 1.363.818 variantes por indivíduo. O software *Admixture* foi utilizado em modo não-supervisionado para a inferência da ancestralidade definindo 2 a 10 componentes de ancestralidade ($k = 2-10$). Foram realizadas análises de *cross-validation* para definir o k com maior precisão preditiva em relação aos dados para as análises, sendo definido aquele de *cross-validation* de menor valor, neste caso $k = 9$ ($cv = 0,43401$; Figura 4). O software gerou um *output* com as porcentagens dos componentes de cada indivíduo; a partir deste *output*, utilizando a função *mean* na Plataforma R (www.r-project.org), foram calculadas as médias de ancestralidade de cada população (https://github.com/ManochioCaique/Genomic-ancestry-and-Star-allele/blob/main/calculo_media_ancestralidade.R).

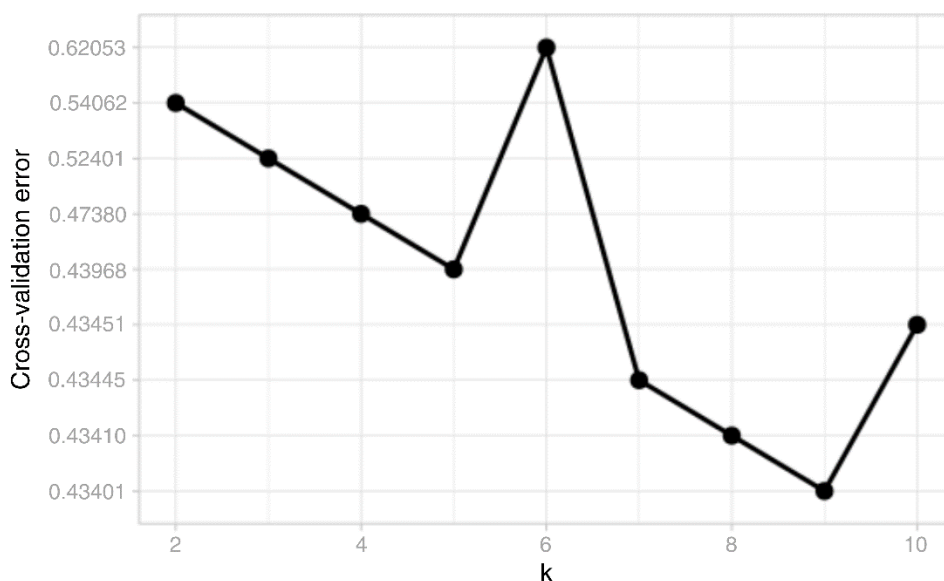


Figura 4: *Cross-validation* plot com valores de CV de $k = 2-10$. Fonte: Do Autor, 2024.

Os dados do Projeto *1000 Genomes* foram baixados para o servidor do Laboratório de Diversidade Genética Humana (LDGH) da Universidade Federal de Minas Gerais (UFMG). Para fins de controle de qualidade foram filtrados, e deste modo, houve a remoção do chr0, variantes A|T e C|G, variantes em heterozigose e dados duplicados e faltantes. O processo foi realizado através do script MosaiQC (<https://github.com/ldgh/MosaiQC-public>). Os dados do sequenciamento do genoma completo de 150 Nativo-americanos do Peru realizado por Harris et al. (2018) foram acessados através de uma colaboração com os autores.

A partir da ferramenta *vcftools* (DANECEK et al., 2011), o gene *ESR1* foi filtrado do arquivo contendo o genoma completo de acordo com a localização descrita pelo *National Center for Biotechnology Information* (NCBI) (151977807..152450754). Os arquivos VCF resultantes continham 13.752 variantes (*1000 Genomes*) e 2.281 variantes (dados originais dos nativos). O cálculo das frequências alélicas foi realizado na plataforma R, utilizando a função *makefreq* do pacote *adegenet* (JOMBART; AHMED, 2011).

O teste de regressão linear e de correlação de Pearson entre as frequências alélicas e a ancestralidade subcontinental ($k = 9$) foi realizado na plataforma R através da função *lm()*. Os scripts utilizados estão disponíveis na página <https://github.com/marianascudeler>. As análises foram realizadas com base na metodologia de Rodrigues-Soares e colaboradores (2020) e, assim como os autores,

consideramos como significativos os valores de $p < 0,01$ (RODRIGUES-SOARES et al., 2020).

As variantes foram anotadas através do software *Annovar* (WANG; LI; HAKONARSON, 2010), utilizando os bancos de dados refGene, clinvar_20221231 e avsnp150 e foram discutidas e analisadas aquelas que possuíam associação descrita com dados clínicos do ClinVar e PharmGKB.

O fluxograma referente aos passos envolvidos nas análises está disponível na Figura 5.

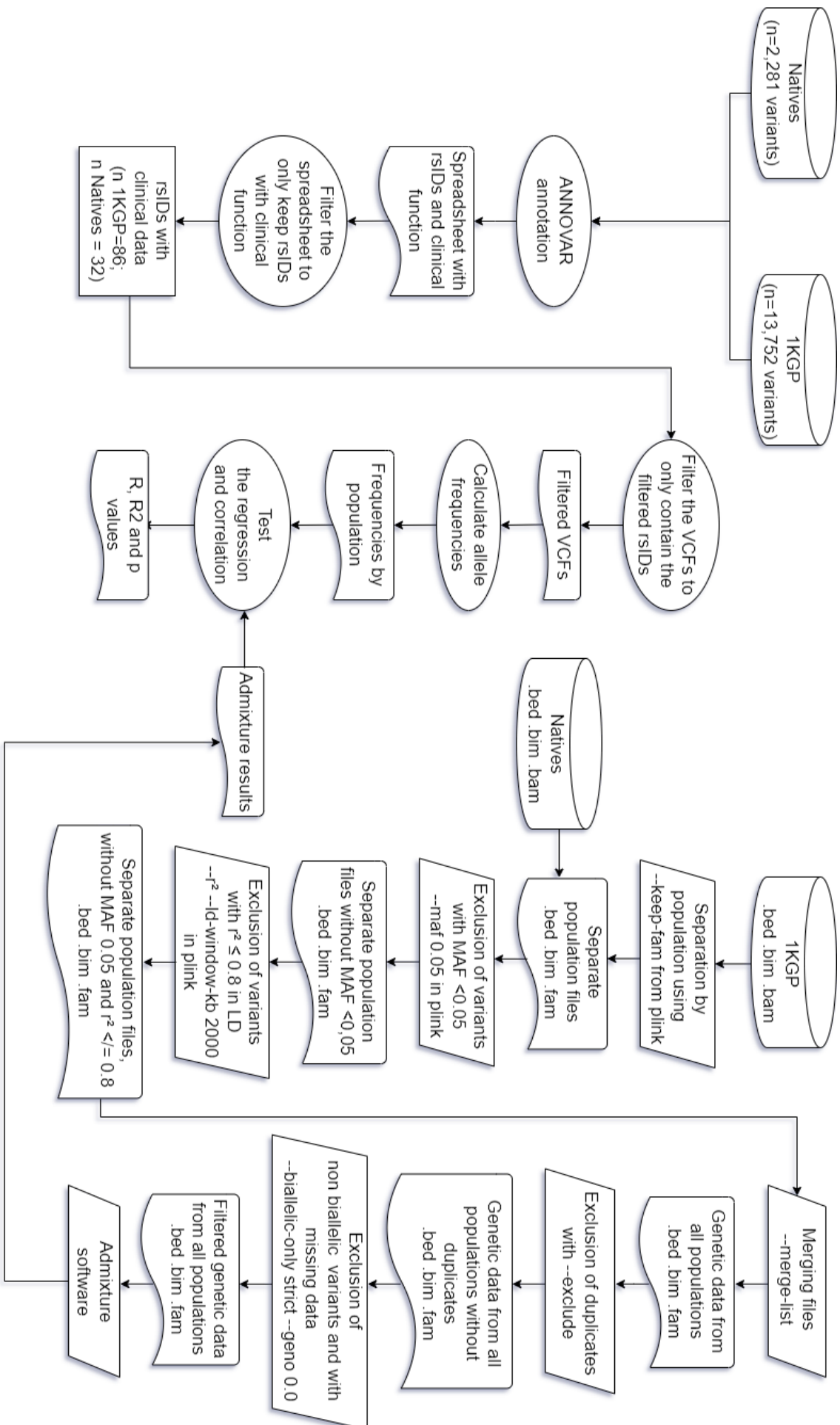


Figura 5: Fluxograma com os passos da metodologia envolvida na realização do estudo. Fonte: Do Autor, 2024.

7. Resultados e Discussão

7.1 Ancestralidade

Considerando a inferência de ancestralidade com $k = 9$ realizada para os 2654 indivíduos oriundos de 26 populações ao redor do mundo do Projeto *1000 Genomes* e as 7 populações nativas, foi possível observar separação subcontinental entre os componentes definidos (Tabela 2).

Tabela 2. Porcentagem de cada componente/ancestralidade para as 33 populações incluídas¹.

POP	NAT2	E-AFR	E-EAS	SAS	NAT	N-EUR	W-AFR	S-EUR
LWK	0.00873	0.820313	0.013251	0.01814	0.008313	0.059826	0.032075	0.029362
ESN	0.008939	0.636553	0.008988	0.007481	0.008701	0.039154	0.265536	0.01685
YRI	0.009532	0.606398	0.009636	0.007655	0.008882	0.038012	0.294888	0.017847
MSL	0.008533	0.426134	0.008774	0.00962	0.009957	0.037669	0.476324	0.015609
GWD	0.009435	0.308329	0.011843	0.0127	0.007334	0.027705	0.577248	0.035814
ACB	0.011873	0.51226	0.012082	0.014665	0.008687	0.062751	0.288676	0.079313
ASW	0.014039	0.452191	0.013592	0.015979	0.034106	0.095907	0.246907	0.113731
CLM	0.043458	0.067437	0.032031	0.0207	0.219678	0.064537	0.047226	0.486352
MXL	0.081647	0.028383	0.056656	0.025129	0.355808	0.045727	0.04188	0.351911
PUR	0.028586	0.093505	0.030807	0.022528	0.124568	0.066168	0.075872	0.539437
PEL	0.147141	0.023475	0.019101	0.01081	0.579486	0.032085	0.020963	0.149875
TSI	0.013438	0.037259	0.024116	0.052179	0.015224	0.107411	0.028541	0.698355
IBS	0.014431	0.03825	0.023308	0.020422	0.015529	0.155306	0.035073	0.673939
GBR	0.019572	0.030868	0.012552	0.024936	0.014377	0.325347	0.027144	0.528343
CEU	0.018553	0.032375	0.011704	0.026211	0.014977	0.333788	0.027952	0.520189
FIN	0.021248	0.027067	0.032186	0.012246	0.009308	0.652192	0.039565	0.195673
PJL	0.020969	0.041763	0.026267	0.582655	0.022779	0.115862	0.032297	0.137857
GIH	0.022429	0.039542	0.01826	0.649843	0.020735	0.103504	0.031867	0.09851
ITU	0.02525	0.037316	0.025763	0.695444	0.017376	0.058443	0.039471	0.071661
STU	0.023182	0.037379	0.030384	0.690256	0.01835	0.057406	0.042368	0.067428
BEB	0.021172	0.035323	0.047955	0.608196	0.024238	0.075879	0.036916	0.063645
CDX	0.017391	0.027565	0.050957	0.023258	0.014563	0.035647	0.024019	0.035535
KHV	0.021736	0.0216	0.110312	0.038392	0.011701	0.028464	0.027813	0.042421
CHS	0.0171	0.0231	0.317831	0.00848	0.012568	0.035338	0.021315	0.035821
CHB	0.018514	0.021376	0.438297	0.009207	0.015172	0.038825	0.02284	0.038769
JPT	0.015989	0.021463	0.712531	0.024579	0.016141	0.036241	0.021573	0.033956
CHOPCCAS	0.221403	0.003379	0.003782	0.002546	0.730057	0.019951	0.003767	0.009861
CUSCO	0.240921	0.019865	0.016962	0.011135	0.534574	0.025957	0.010092	0.131493
IQUITOS	0.087343	0.00756	0.036657	0.006909	0.676822	0.009515	0.018539	0.143418

MATZES	0.020816	0.003496	0.031472	0.00537	0.912193	0.002105	0.00403	0.008856
MOCHES	0.025747	0.008796	0.007123	0.006429	0.87903	0.039441	0.004255	0.018435
TRUJILLO	0.098724	0.021439	0.017725	0.013958	0.702291	0.025725	0.011935	0.095919
UROS	0.850807	0.000662	0.00198	0.001726	0.135974	0.003403	0.000579	0.004132

¹N-EUR: Norte-europeu; W-AFR: Oeste-africano; NAT: Nativo-americana 1; NAT2: Nativo-americana 2
W-EAS: Leste-asiático - Oeste; S-EUR: Sul-europeu; E-AFR: Leste-africano; E-EAS: Leste-asiático - Leste; SAS: Sul-asiático; POP: população; LWK: Luhya do Quênia; ESN: Esan da Nigéria; YRI: Yoruba da Nigéria; MSL: Mende de Serra Leoa; GWD: Mandika de Gâmbia; ACB: Afro-Caribenhos de Barbados; ASW: Indivíduos com ancestralidade africana no sudeste dos EUA; CLM: Colombianos de Medellín na Colômbia; MXL: Indivíduos com ancestralidade Mexicana vivendo em Los Angeles; PUR: Porto-Riquenhos de Porto Rico; PEL: Peruanos de Lima; TSI: Italianos da Toscana; IBS: Ibéricos da Espanha; GBR: Britânicos da Inglaterra e Escócia; CEU: Descendentes de europeus de Utah; FIN: Finlandeses da Finlândia; PJL: Paquistaneses; GIH: Indianos Gujarati em Houston EUA; ITU: Indianos Telugu no Reino Unido; STU: Sri Lanka de Tamil no Reino Unido; BEB: Bengali de Bangladesh; CDX: Chineses Dai de Xishuangbanna; KHV: Vietnamitas Kinh da cidade Ho Chi Minh; CHS: Chineses Han do Sul; CHB: Chineses Han de Pequim; JPT: Japoneses de Tóquio. Fonte: Do Autor, 2024.

Mais especificamente, as diferenciações subcontinentais foram observadas com o estabelecimento de três componentes asiáticos (W-EAS, W-EAS e SAS), dois componentes africanos (W-AFR e E-AFR), dois componentes europeus (N-EUR e S-EUR) e, por fim, dois componentes nativo-americanos (NAT-1 e NAT-2) (Figura 6). Diferenciações subcontinentais também foram observadas em múltiplos estudos na literatura, especialmente considerando as populações europeias, africanas e asiáticas (AUTON et al., 2015; GOUVEIA et al., 2023; HOMBURGER et al., 2015; KEHDY et al., 2015; LAZARIDIS et al., 2014; PAGANI et al., 2016; WANG et al., 2021).

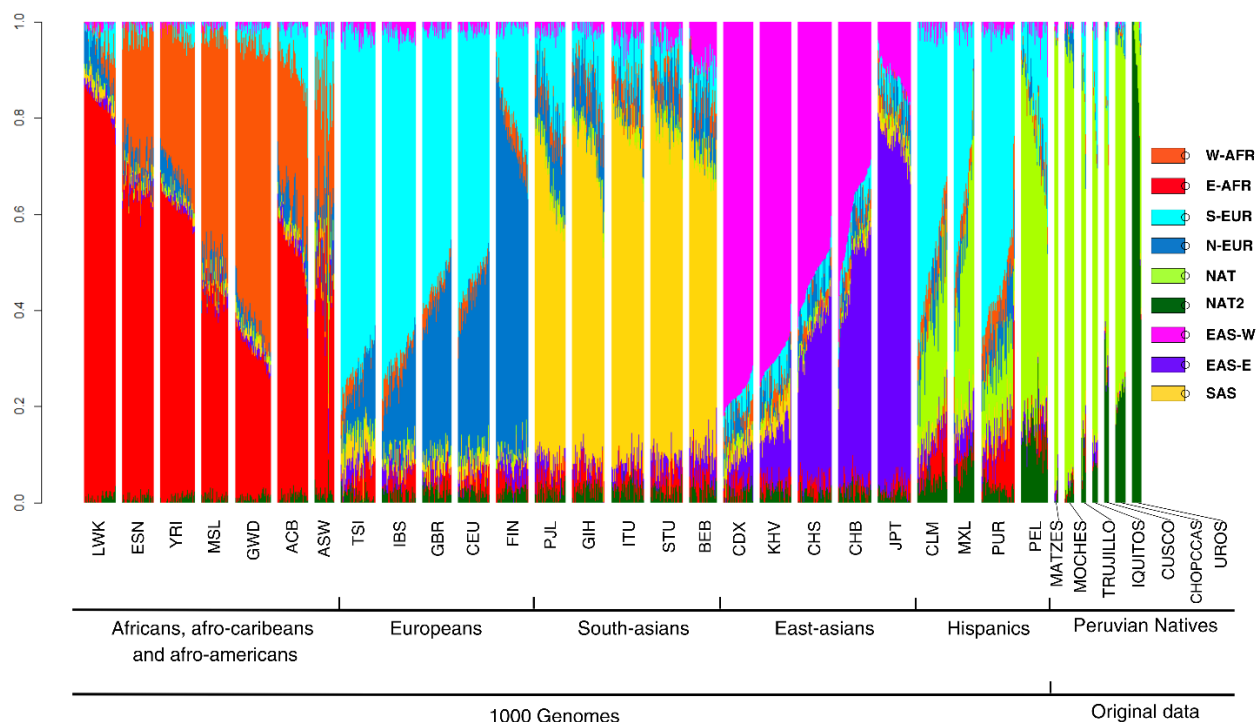


Figura 6. Gráfico de barras da ancestralidade por população. W-AFR: Oeste da África; E-AFR: Leste da África; S-EUR: Sul da Europa; N-EUR: Norte da Europa; NAT: Nativo-americano; NAT2: Nativo-americano 2 SAS: Sul Asiático; W-EAS: Leste Asiático ao Oeste; E-EAS: Leste Asiático ao Leste; LWK: Luhya do Quênia; ESN: Esan da Nigéria; YRI: Yoruba da Nigéria; MSL: Mende de Serra Leoa; GWD: Mandinka de Gâmbia; ACB: Afro-Caribenhos de Barbados; ASW: Indivíduos com ancestralidade africana no sudeste dos EUA; CLM: Colombianos de Medellín na Colômbia; MXL: Indivíduos com ancestralidade Mexicana vivendo em Los Angeles; PUR: Porto-Riquenhos de Porto Rico; PEL: Peruanos de Lima; TSI: Italianos da Toscana; IBS: Ibéricos da Espanha; GBR: Britânicos da Inglaterra e Escócia; CEU: Descendentes de europeus de Utah; FIN: Finlandeses da Finlândia; PJL: Paquistaneses; GIH: Indianos Gujarati em Houston EUA; ITU: Indianos Telugu no Reino Unido; STU: Sri Lanka de Tamil no Reino Unido; BEB: Bengali de Bangladesh; CDX: Chineses Dai de Xishuangbanna; KHV: Vietnamitas Kinh da cidade Ho Chi Minh; CHS: Chineses Han do Sul; CHB: Chineses Han de Pequim; JPT: Japoneses de Tóquio. Fonte: Do Autor, 2024.

Na Figura 6, temos populações africanas com dois componentes subcontinentais, W-AFR (oeste africano) e E-AFR (leste africano). Nota-se uma clara e gradual diminuição da proporção do componente E-AFR à medida em que se vai do leste a oeste do continente, passando pela população do Quênia (82%), Nigéria (63% - ESN; 60% - YRI), Serra Leoa (42%) e Gâmbia (30%). Auton e colaboradores (2015) realizaram análise semelhante incluindo as mesmas populações (Projeto *1000 Genomes*) e definindo k entre 5 e 12 utilizando um menor número de SNPs. Os resultados encontrados para $k = 9$ são bastante similares, com exceção da separação observada para a população de Gâmbia (GWD). No entanto, pode-se perceber que a partir do $k = 10$ as populações africanas são diferenciadas em 3 componentes subcontinentais, onde as populações do Quênia (LWK), Nigéria (ESN e YRI) e Gâmbia (GWD) estão bem diferenciadas uma das outras, cada uma com um componente; a população da Serra Leoa exibe os componentes nigeriano e gambiano, este em maior proporção, o que é coerente dada a localização geográfica dos países (Figura 7) (AUTON et al., 2015).

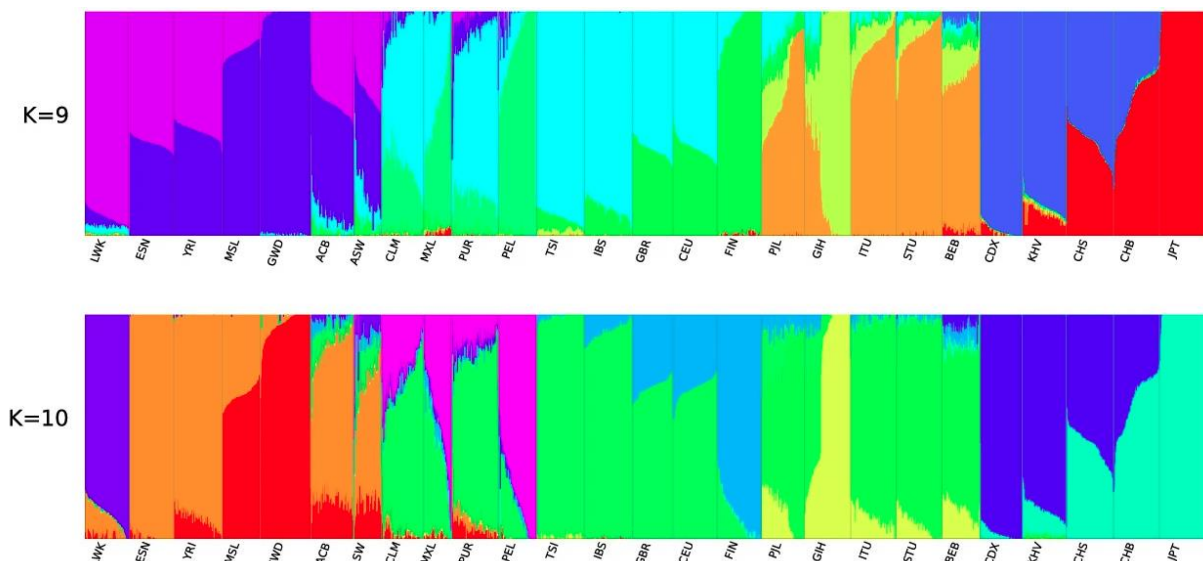


Figura 7: Ancestralidade com $k = 9-10$ dos indivíduos do Projeto *1000 Genomes* no trabalho de Auton e colaboradores (2015). Fonte: Adaptado de (AUTON et al., 2015).

Gouveia e colaboradores analisaram 6.267 indivíduos com mais de 10% de ancestralidade africana de 25 populações: 9 populações miscigenadas da América, 11 africanas, 2 europeias e 3 nativo-americanas; os dados são oriundos do projetos EPIGEN-Brasil, projeto *1000 Genomes*, *Prostate*, *Lung*, *Colorectal and Ovarian Cancer screening*, *Ghana Prostate Study Accra*, *Epidemiology of Burkitt's Lymphoma*, além do *Instituto Nacional de Salud* do Peru e *Tishkoff Laboratory*; os autores estabeleceram 4 componentes africanos: Oeste-central, sul/oriental, ocidental e norte da Uganda (GOUVEIA et al., 2020). Diferentemente de nossas análises e Gouveia e colaboradores (2020), alguns estudos não observaram miscigenação de 2 componentes na população Yoruba (HOMBURGER et al., 2015; KEHDY et al., 2015; LAZARIDIS et al., 2014; PAGANI et al., 2016; TISHKOFF et al., 2009; WANG et al., 2021), o que pode se dever ao set de SNPs, maior k e/ou número de indivíduos e de populações estudadas. Além disso, é possível observar que alguns autores não observaram a separação entre as populações Luhya e Yoruba constatada em nossas análises (LAZARIDIS et al., 2014; PAGANI et al., 2016); Gouveia e colaboradores e Kehdy e colaboradores, no entanto, obtiveram resultados semelhantes aos nossos ($k = 6$ e $k = 9-15$, respectivamente) (GOUVEIA et al., 2020; KEHDY et al., 2015).

Em seguida temos as populações miscigenadas de origem africana, os afro-caribenhos (ACB) e afro-americanos dos Estados Unidos (ASW). Considerando estas populações, observa-se proporções semelhantes de ambos os componentes africanos. Em suas análises, Gouveia et al. observaram proporções bastante semelhantes no que diz respeito aos componentes ancestrais de ACB e ASW, com exceção da presença de um terceiro componente africano (sul/oriental) (GOUVEIA et al., 2020).

Com relação às populações europeias, a separação subcontinental observada entre norte e sul está de acordo com o que é relatado na literatura por outros estudos e reforça a constatação de que é necessário começar a considerar as populações europeias como miscigenadas em nível subcontinental para que se evitem possíveis constatações espúrias em estudos genéticos (AUTON et al., 2015; GOUVEIA et al., 2023; HOMBURGER et al., 2015). Assim como em relação às populações africanas, é notável a clara e gradual diminuição da proporção do componente S-EUR (sul-europeu) entre as populações à

medida em que se vai do sul ao norte do continente (69% nos italianos, 67% nos espanhóis, 52% nos ingleses e escoceses e 19% nos finlandeses); o mesmo foi observado por Homburguer e colaboradores (HOMBURGER et al., 2015). Um estudo que teve como objetivo avaliar a ancestralidade subcontinental europeia, por sua vez, avaliou a presença de um terceiro componente, descrevendo o norte-europeu, sudeste-europeu e sudoeste-europeu. Os autores também descrevem uma clara separação da população finlandesa, como aqui foi descrito (GOUVEIA et al., 2023). A população CEU, que é formada por indivíduos descendentes de europeus residentes em Utah, possui proporções bastante semelhantes às observadas nos ingleses e escoceses [52% vs. 52% para S-EUR e 33% vs. 32% para N-EUR (norte-europeu), respectivamente], o que vai de acordo com a história de formação da população estadunidense, cujo território foi colonizado pelos ingleses (TSUKADA, 2021), e como também foi descrito por Gouveia e colaboradores e Auton e colaboradores (AUTON et al., 2015; GOUVEIA et al., 2023).

Dentre as populações asiáticas, foram observadas diferenças subcontinentais que separaram os componentes em 3: um sul-asiático e dois leste-asiáticos. As populações sul-asiáticas se mostraram bem diferenciadas do leste-asiático e bastante semelhantes entre si, o que é similar ao que foi observado por Pagani e colaboradores em suas análises (PAGANI et al., 2016). Dentre as populações leste-asiáticas, a população japonesa se mostrou diferenciada das chinesas (71% do componente E-EAS – leste do leste-asiático), e dentre as chinesas a diferenciação é estabelecida e evidenciada até mesmo entre etnias (0,5% vs. 31% e 43%, respectivamente, para as populações Dai e Han do Sul e de Pequim quanto ao E-EAS), motivo pelo qual os componentes leste-asiáticos foram divididos em Leste e Oeste. Lazaridis e colaboradores e Tishkoff e colaboradores, no entanto, não observaram diferenças tão consideráveis entre as etnias chinesas e japoneses (LAZARIDIS et al., 2014; TISHKOFF et al., 2009). Em contraste, um estudo que buscou avaliar a estrutura genética e ancestralidade de populações chinesas, coreanas e japonesas concluiu que a composição genética de tais populações pode ser muito bem distinguida tanto por abordagens de genoma completo quanto por marcadores informativos de ancestralidade (AIMs). As diferenças até mesmo entre as etnias Dai e Han também foram estabelecidas, assim como nas análises aqui realizadas (WANG et al., 2018).

Finalizando as populações do Projeto *1000 Genomes*, dentre os Hispânicos, estão incluídos colombianos (CLM), mexicanos residentes nos Estados Unidos (MXL), porto-riquenhos (PUR) e peruanos (PEL). Para todas essas populações é possível observar a

presença de um maior número de componentes (ou seja, ancestralidades) em comparação às outras populações, o que é coerente com o fato de que se trata de populações miscigenadas de origens ancestrais semelhantes (nativo-americanos, europeus e africanos) (HOMBURGER et al., 2015). Além disso, destaca-se a alta proporção de ancestralidade nativa na população peruana. Um trabalho que teve como objetivo avaliar a história demográfica e ancestralidade da América do Sul relatou que a população peruana tem a maior proporção de ancestralidade Nativo-americana dentre as populações sul-americanas, especificamente Nativo-americanas dos Andes (HOMBURGER et al., 2015). Nossos dados, no entanto, mostram uma menor proporção de ancestralidade dos Andes nos peruanos do *1000 Genomes* (Cusco, Chopccas e Uros – componente NAT2 – nativo-americano 2). A população de Lima analisada por Harris e colaboradores (2018) revelou-se com proporções bastante semelhantes de ancestralidade nativa com a presença de um componente europeu.

Os resultados das 7 populações restantes foram gerados com dados originais de Nativos do Peru (Matzes, Moches, Chopccas, Uros, Trujillo, Iquitos e Cusco). Como para os africanos, europeus e asiáticos, também pudemos observar a separação de componentes subcontinentais para estas populações, os componentes nativos (NAT – nativo-americano e NAT2). A população Moches, da costa do Peru, se mostrou muito diferenciada dos Uros, nos Andes. Pode-se visualizar o componente NAT, que caracteriza os Moches, diminuindo em proporção em relação ao componente NAT2 à medida que se vai da costa, se passa pela região amazônica e chega-se aos Andes. As populações e suas regiões específicas podem ser visualizadas na Figura 8. No ano de 2018, Harris e colaboradores publicaram na revista *Proceedings of the National Academy of Sciences* (PNAS), uma das mais relevantes revistas científicas multidisciplinares do mundo, o maior painel com dados de sequenciamento de genoma completo de Nativos, que incluiu os 150 genomas aqui também analisados em colaboração com os autores, juntamente a dados de *array* de genotipagem de mais 130 indivíduos indígenas e/ou miscigenados com alta ancestralidade indígena com o objetivo de avaliar a dinâmica genômica evolutiva. Os autores fizeram análises de ancestralidade de $k = 4$ a 8. Os resultados do $k = 7$ são bastante semelhantes aos que nós encontramos em nossas análises principalmente no que diz respeito à diferenciação entre as populações Moches e Uros, no entanto o mesmo não é observado para $k = 4, 5, 6$ e 8. Isto pode se dever a diferenças no set de SNPs selecionados para realizar as análises: enquanto neste trabalho foram incluídas 1.363.818 variantes,

Harris e colaboradores (2018) incluíram 183.579. Resultados encontrados por Homburger e colaboradores (2015), no entanto, mostram clara diferenciação entre populações indígenas dos Andes e populações não-andinas (o set dos autores incluía 364.470 SNPs).

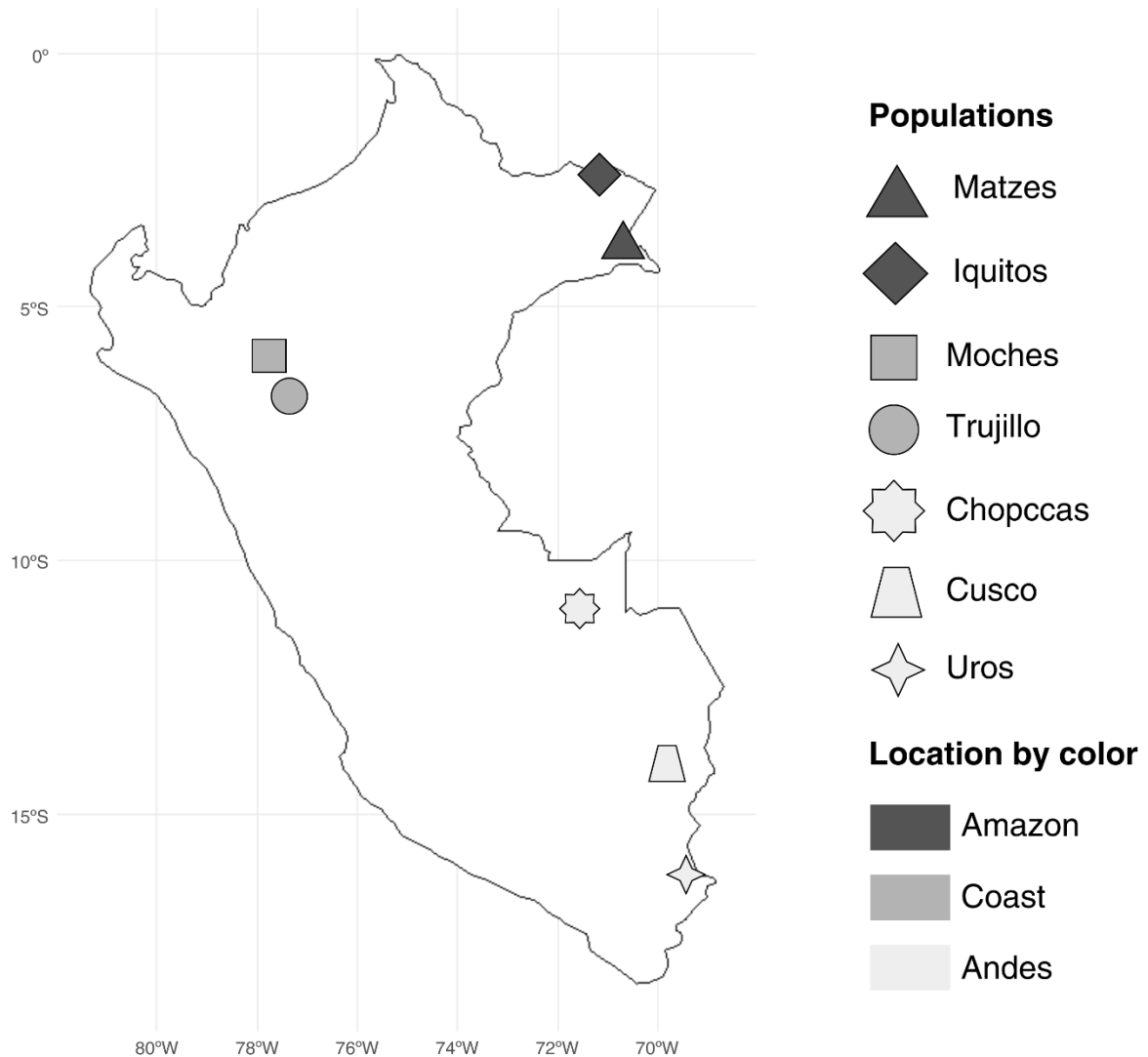


Figura 8. Representação das localizações das populações Nativas peruanas incluídas no estudo. Fonte: Do Autor, 2024.

7.2 Anotação das variantes

No total, os arquivos VCF do Projeto *1000 Genomes* e dos Nativos continham 13.752 e 2.281 variantes, respectivamente. A partir da anotação dos dados pelo *software* ANNOVAR (WANG; LI; HAKONARSON, 2010), as variantes foram filtradas em acordo com a disponibilidade de dados clínicos no ClinVar e na aba Clinical Annotations do PharmGKB (Figura 9). Com isso, foram selecionadas 86 variantes do Projeto *1000 Genomes* e 32 dos dados de Nativos para que suas frequências fossem calculadas e se

testasse a regressão e correlação entre as frequências dessas variantes e os 9 componentes de ancestralidade. Das 32 variantes dos dados de Nativos selecionadas, apenas o rs747405346 não está presente nos dados do *1000 Genomes*.

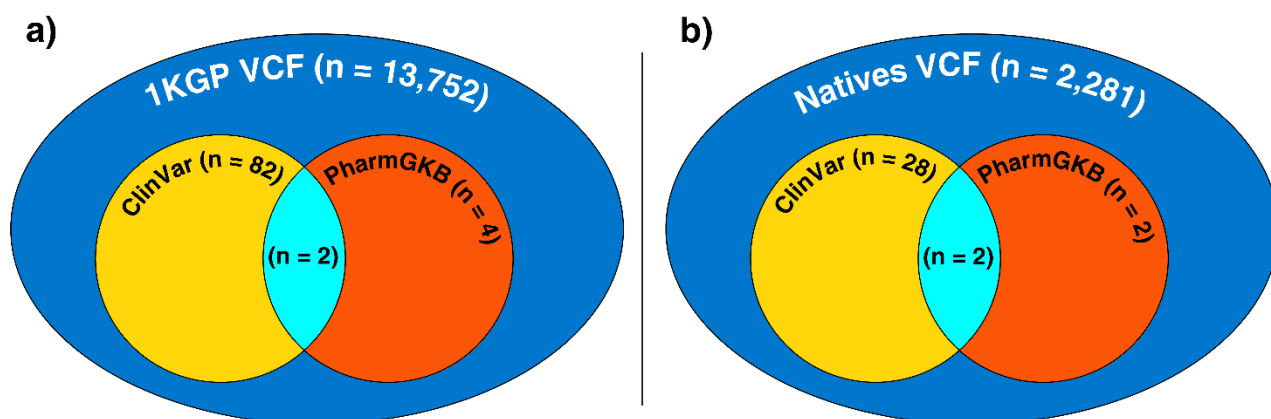


Figura 9. Diagrama de Venn representando o número de variantes em acordo com os critérios de seleção para análises de regressão e associação (a) Variantes do Projeto *1000 Genomes*; (b) Variantes dos dados originais de Nativos peruanos. Fonte: Do Autor, 2024.

Através da anotação foi também possível identificar que cerca de metade ($n = 42$ – 48,8% e 16 – 50% para o Projeto *1000 Genomes* e os dados de Nativos peruanos, respectivamente) das variantes analisadas está localizada em regiões intrônicas, em contraste a 34,8% e 28,1% variantes em éxons ($n = 30$ e 9). Neste contexto é importante destacar o trabalho de Yang e colaboradores (2021), em que os autores agregaram os dados de variantes farmacogenéticas de farmacogenes de 4 bancos de dados da área: *Drug Bank*, *PharmGKB*, *PharmaADME* e *Biotransformation genes*. Através das análises, estabeleceram que 48% ($n = 1566$) das variantes estava em regiões intrônicas, reforçando assim como os resultados aqui encontrados a importância de se estudar genomas completos em comparação a exomas (YANG et al., 2021).

No que diz respeito ao banco de dados refGene, das 86 variantes incluídas do Projeto 1000 Genomas, 66 foram listadas como possivelmente associadas à resistência ao estrogênio e 19 com distrofia muscular de Emery-Dreifuss e ataxia espinocerebelar. Em relação às variantes dos dados de Nativos, por sua vez, 24 das 32 variantes foram listadas como possivelmente associadas à resistência ao estrogênio e 8 com distrofia muscular de Emery-Dreifuss e ataxia espinocerebelar.

No entanto, é importante destacar que se notou uma limitação da anotação que deve ser levada em consideração em estudos posteriores e no desenho de análises de anotação. De acordo com o NCBI em relação ao genoma de referência 37, existe uma determinada região de sobreposição entre os genes *ESR1* e *SYNE1* presentes em fitas opostas. Levando em conta que um polimorfismo sempre ocorrerá em um mesmo *locus* nas duas fitas de DNA, é possível que um polimorfismo do *ESR1* seja identificado como do *SYNE1* caso a análise não especifique a fita analisada, e vice-versa. Com isso, observamos que das 86 e 32 variantes filtradas para nossas análises, 19 e 8 foram anotadas como variantes do gene *SYNE1*. É importante que, para aperfeiçoar os processos de anotação futuramente, a limitação causada pela sobreposição de genes seja considerada afim de máxima acurácia da anotação.

7.3 Associação entre ancestralidade e frequências alélicas e anotação de variantes

A partir da análise de regressão linear e correlação de Pearson realizada entre as frequências alélicas de variantes do *ESR1* e os 9 componentes de ancestralidade (W-AFR, E-AFR, S-EUR, N-EUR, W-EAS, E-EAS, SAS, NAT e NAT2), observamos, no total, 145 associações positivas e negativas consideradas significativas ($p < 0,01$) presentes em 73 variantes (APÊNDICE A). Destas, 106 são positivas (em 70 variantes), indicando que quanto maior a proporção da determinada ancestralidade maiores serão as chances do indivíduo ser carreador de tais variantes, e 39 negativas (28 variantes), indicando justamente o contrário: quanto maior a proporção de ancestralidade, menores as chances de se ter a variante (73,1% e 26,8% das associações, respectivamente).

A respeito da associação aos 9 componentes, destaca-se que 53,1% das associações foram encontradas com os componentes de ancestralidade africana (W-AFR e E-AFR) (Figura 10). Considerada a quantidade substancial de associações e o papel do *ESR1* em vias de alta relevância no contexto do tratamento hormonal de câncer de mama, um ponto que chama atenção é o fato de que mulheres com ancestralidade africana possuem maior risco de mortalidade em decorrência da doença (NEWMAN, 2014). Levando em conta a subrepresentação de populações africanas em estudos genômicos e em estudos farmacogenéticos do câncer de mama, reforça-se a chamada para a inclusão dessas populações em estudos genéticos, uma vez que dessa maneira pode ser possível

encontrar meios de se preencher lacunas acerca desta questão no contexto que compete à genética e farmacogenética (SCUDELER et al., 2023; SIRUGO; WILLIAMS; TISHKOFF, 2019).

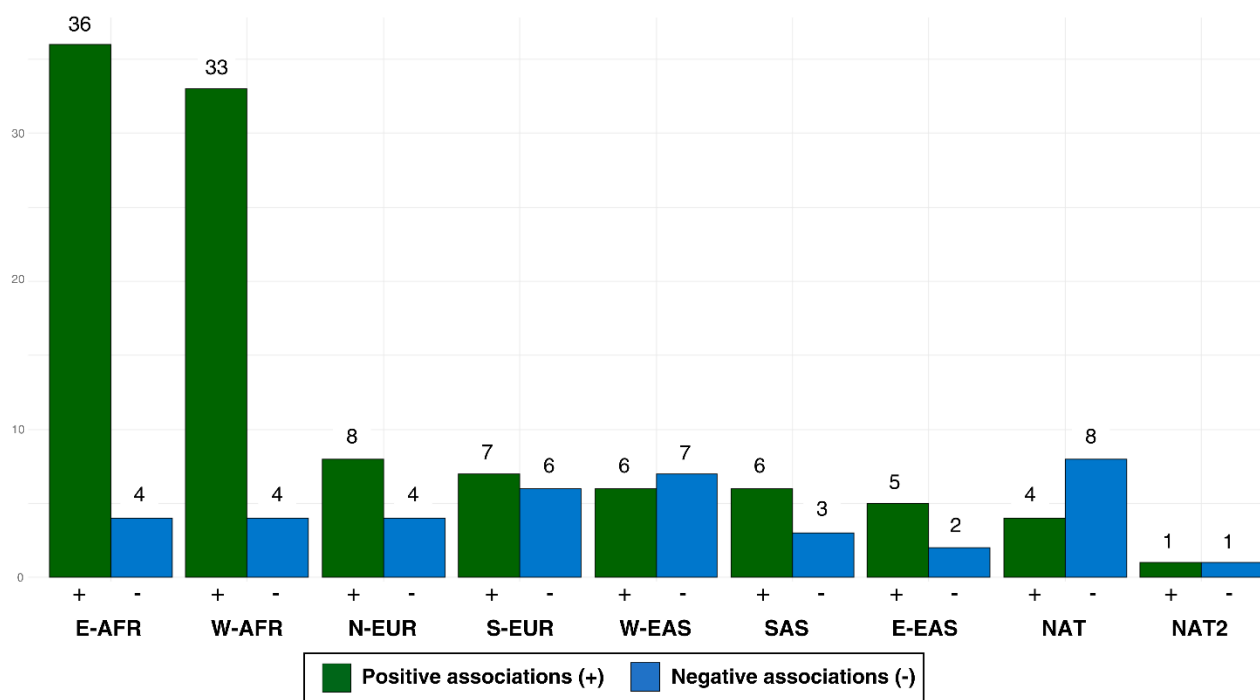


Figura 10. Número de associações positivas e negativas por componente de ancestralidade. N-EUR: Norte-europeu; W-AFR: Oeste-africano; NAT: Nativo-americana; NAT2: Nativo-americana 2; W-EAS: Leste-asiático - Oeste; S-EUR: Sul-europeu; E-AFR: Leste-africano; E-EAS: Leste-asiático - Leste; SAS: Sul-asiático. Fonte: Do Autor, 2024.

Buscando avaliar com mais profundidade a extensão das diferenças subcontinentais entre as associações de frequências alélicas de variantes do *ESR1* e os componentes de ancestralidade genômica, observou-se que, considerando as variantes por *cluster* de continente (i.e., avaliando SAS + W-EAS + E-EAS, W-AFR + E-AFR, N-EUR + S-EUR e NAT + NAT2), havia variantes exclusivas do componente subcontinental ao qual foram associadas considerando seus outros componentes continentais (APÊNDICE A). É importante mencionar que quando se diz “exclusivas” significa que a variante foi associada com apenas um dos componentes do *cluster* continental. Ou seja, variantes cujas frequências foram associadas com o componente subcontinental SAS e não a W-EAS e E-EAS, e assim subsequentemente em relação aos outros *clusters*. Quando se diz “não exclusivas”, por sua vez, entende-se que a variante foi associada a todos os componentes do *cluster* continental. É possível observar que a grande maioria das variantes associadas à ancestralidade nativa (NAT) são exclusivas a ela em relação a

NAT2, sugerindo uma diferenciação subcontinental considerável na distribuição de variantes dentre as populações nativas, e o mesmo é observado para os componentes asiáticos e europeus (Tabela 3). Isso sugere a importância do papel das diferenças subcontinentais no contexto da implementação Farmacogenética, incluindo o design de painéis farmacogenéticos, e reforça os perigos de se extrapolar dados genéticos de uma população para outra (BERGSTRÖM et al., 2020; BUSTAMANTE; DE LA VEGA; BURCHARD, 2011; SIRUGO; WILLIAMS; TISHKOFF, 2019). Em contraste, considerando os componentes africanos é possível observar o contrário: um maior número de variantes associadas a ambos os componentes africanos (isto é, não exclusivas), apontando para uma maior similaridade continental.

Tabela 3. Número de variantes exclusivas e não exclusivas por componente de ancestralidade dentre os *clusters* de continentes (SAS + W-EAS + E-EAS; NAT + NAT2; S-EUR + N-EUR; W-AFR + E-AFR).

Componentes asiáticos	Número de variantes associadas
Exclusivas SAS	7
Exclusivas W-EAS	6
Exclusivas E-EAS	2
Não exclusivas	14
Componentes nativos	
Exclusivas NAT2	1
Exclusivas NAT	11
Não exclusivas	2
Componentes europeus	
Exclusivas S-EUR	7
Exclusivas N-EUR	6
Não exclusivas	12
Componentes africanos	
Exclusivas W-AFR	4
Exclusivas E-AFR	7
Não exclusivas	66

Fonte: Do autor (2024). N-EUR: Norte-europeu; W-AFR: Oeste-africano; NAT: Nativo-americana 1; NAT2: Nativo-americana 2 W-EAS: Leste-asiático - Oeste; S-EUR: Sul-europeu; E-AFR: Leste-africano; E-EAS: Leste-asiático - Leste; SAS: Sul-asiático. Fonte: Do Autor, 2024.

7.4 Variantes listadas no ClinVar e PharmGKB

No que diz respeito a variantes anotadas com dados clínicos disponíveis no ClinVar por múltiplas submissões, destacamos que rs12681 (associada positivamente com W-AFR e E-AFR e negativamente com SAS E W-EAS – APÊNDICE A), rs2250122

(associada positivamente com NAT e negativamente com W-AFR e E-AFR), rs200880341 (associada positivamente com E-EAS), rs2295190 (associada positivamente com N-EUR), rs2295191 (associada positivamente com E-EAS e W-EAS) e rs2295192 foram listadas como benignas ou provavelmente benignas em relação a ataxia autossômica recessiva tipo Beauce e distrofia muscular de Emery-Dreifuss. Em relação às mesmas condições, rs567194577 e rs144206837, por sua vez, são consideradas variantes com dados conflitantes de patogenicidade. As variantes rs2077647 e rs2228480 (associada positivamente a NAT) foram listadas como benignas ou provavelmente benignas em relação a susceptibilidade a enxaqueca com/ou sem aura, infarto do miocárdio, síndrome de resistência ao estrogênio e câncer de mama hereditário. Além disso, rs200075329 (associada positivamente a NAT), rs78574022 e rs9341068 (ambas associadas positivamente a W-AFR e E-AFR) foram consideradas benignas ou provavelmente benignas enquanto rs139740651 (associada positivamente a E-AFR) e rs139960913 (associada positivamente a N-EUR), por sua vez, possuem significância ainda incerta e dados conflitantes de patogenicidade, respectivamente. Quanto às variantes anotadas na aba de *Clinical Annotations* do PharmGKB, ressaltamos a variante rs2234693 que, além de ser classificada como fator de risco a infarto do miocárdio pelo ClinVar, está associada ao risco de toxicidade devido ao uso de letrozol e anastrozol para o tratamento de neoplasias e metanfetamina para o tratamento de doenças psíquicas, além de resposta insatisfatória ao tratamento de artrite reumatóide com leflunomida (genótipo CC) (WHIRL-CARRILLO et al., 2012). A variante rs9340799, por sua vez, está associada a risco de toxicidade em resposta ao uso de tamoxifeno, letrozol e anastrozol, e resposta diminuída a leflunomida e maior aumento da densidade mineral óssea da espinha em resposta à reposição hormonal com estrogênio conjugado e medroxiprogesterona (genótipo GG). Associada a toxicidade devido ao tratamento de neoplasias com agentes alquilantes e cisplatina foi a variante rs2207396, que é negativamente associada a NAT e NAT2 e positivamente associada a SAS. Já rs2813543, positivamente associada aos componentes africanos W-AFR e E-AFR e negativamente associada a S-EUR, é listada como associada a toxicidade em resposta ao tratamento de câncer de mama com exemestano, assim como rs9322335 (positivamente associada ao componente S-EUR e negativamente a E-EAS) em relação a exemestano e letrozol e rs4870061 (positivamente associada a ambos os componentes africanos e negativamente associada a ambos os componentes leste-asiáticos) em relação ao letrozol.

Além disso, destacamos as variantes rs9340799 e rs2234693 que, além do citado previamente, foram consideradas as principais variantes do gene *ESR1* para as quais foram encontradas associações com alterações na resposta a tratamentos do câncer de mama em uma revisão sistemática do nosso grupo focada na farmacogenética publicada em 2022 (SCUDELER et al., 2023). Segundo as nossas análises de regressão, rs9340799 é negativamente associada ao componente nativo NAT ($p = <0,0002$; $\beta = -0,1899$), como é possível ver na Figura 11(a) onde as populações nativas de regiões amazônicas e da costa estão bem destacadas das demais, e positivamente associada ao componente sul-asiático ($p = <0,0006$; $\beta = 0,2277$), cujas populações também estão bem destacadas Figura 11(b).

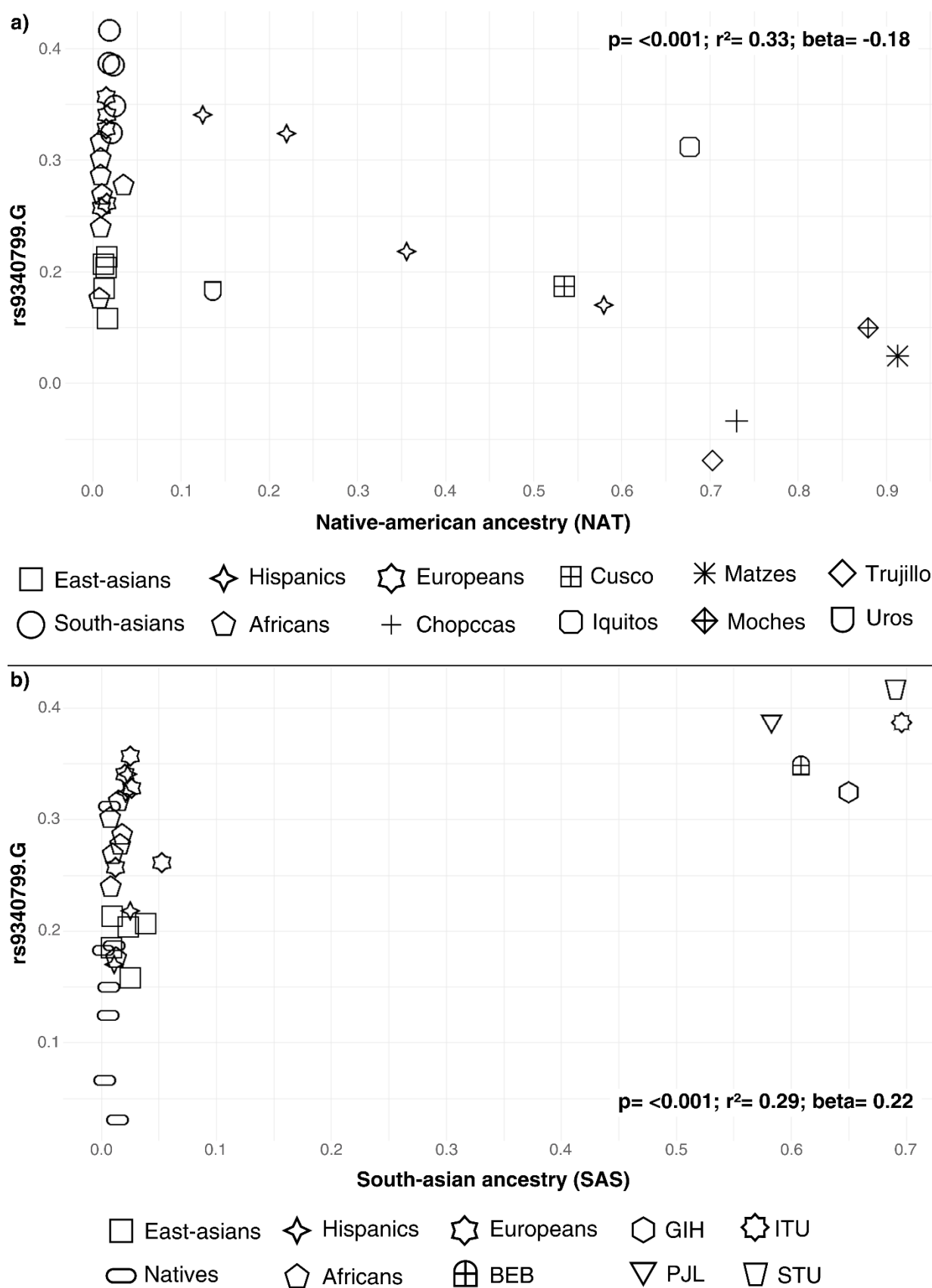
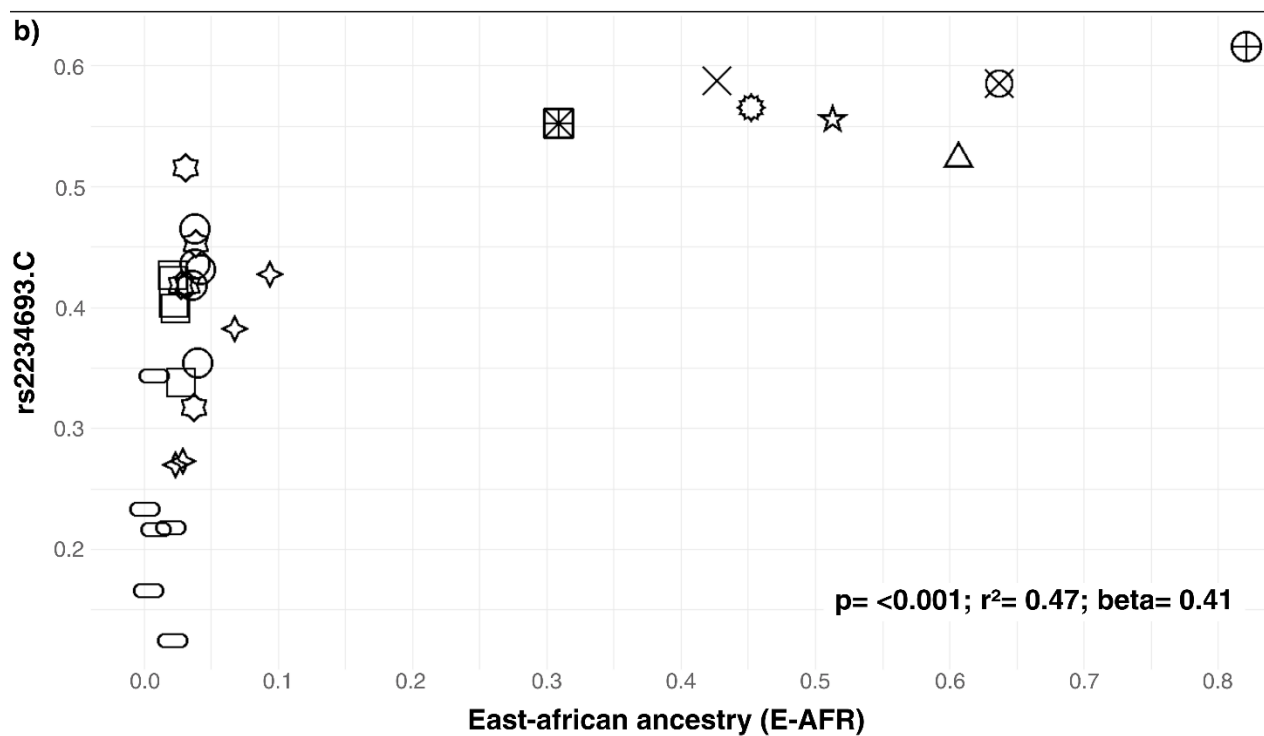
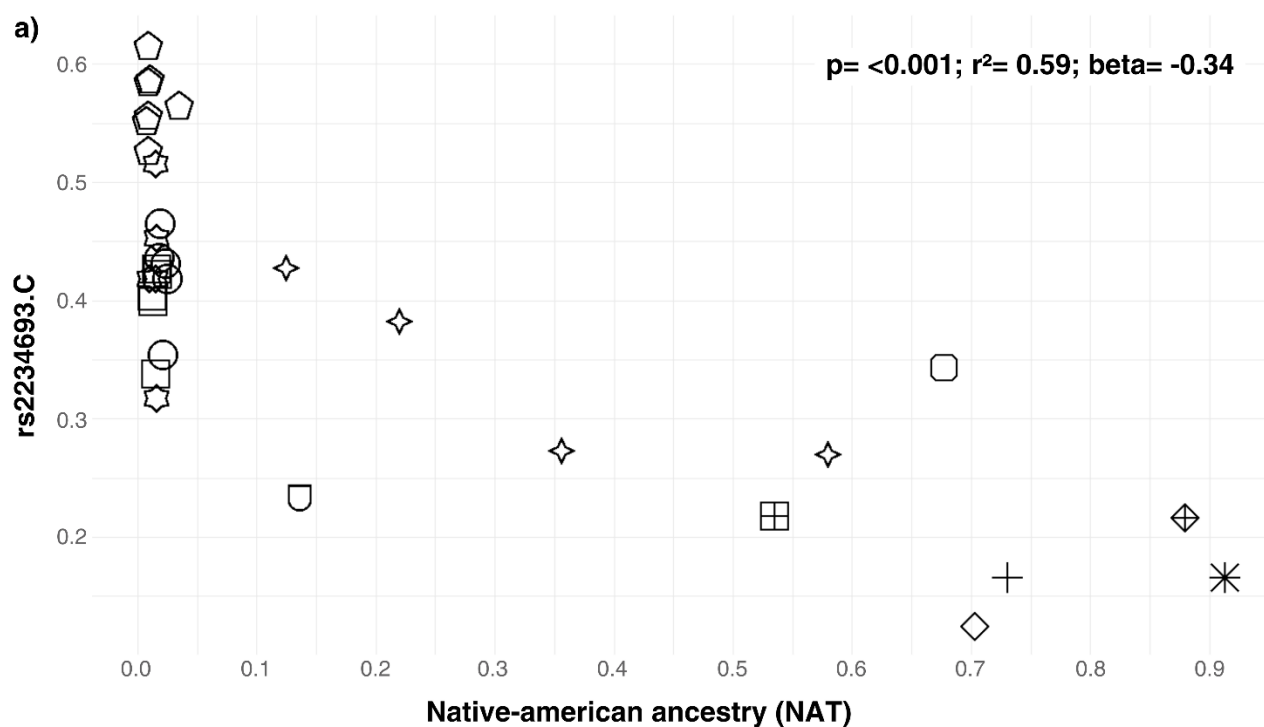


Figura 11. Gráfico de distribuição de frequência alélica (rs9340799.G) em relação à ancestralidade. (a) Ancestralidade Nativa-americana (NAT); (b) Ancestralidade Sul-asiática (SAS). PJL: Paquistaneses; GIH:

Indianos Gujarati em Houston EUA; ITU: Indianos Telugu no Reino Unido; STU: Sri Lanka de Tamil no Reino Unido; BEB: Bengali de Bangladesh. Fonte: Do Autor, 2024. Fonte: Do Autor, 2024.

A variante rs2234693, por sua vez, também é negativamente associada ao componente NAT ($p = <0,0001$; $\beta = -0,3497$), sendo possível observar o mesmo padrão na Figura 12(a), e positivamente associada às ancestralidades africanas E-AFR e W-AFR ($p = <0,0001$; $\beta = 0,4152$; $p = <0,0001$; $\beta = 0,5719$). Neste caso, é possível visualizar as populações africanas bem destacadas das demais na Figura 12(b,c).



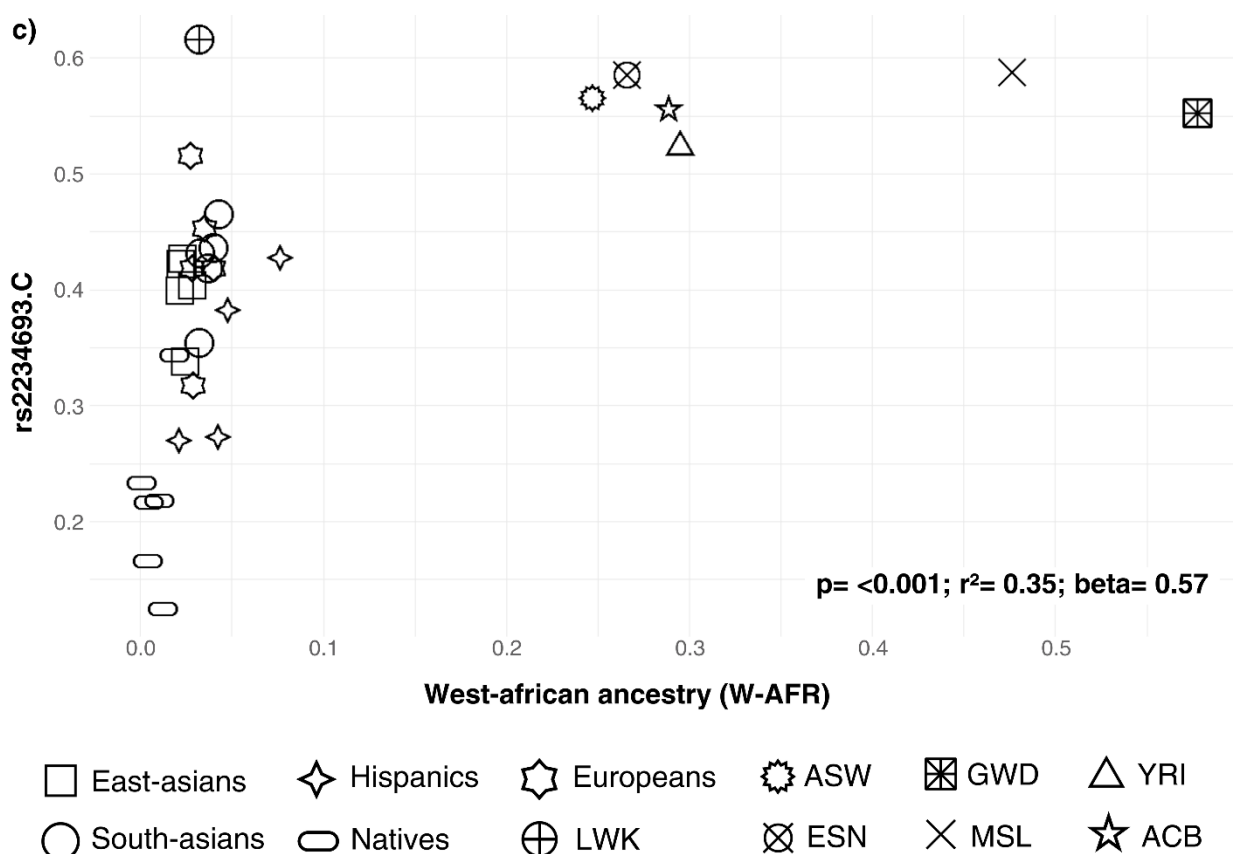


Figura 12. Gráfico de distribuição de frequência alélica (rs9340799.G) em relação à ancestralidade. (a) Ancestralidade Nativa-americana (NAT); (b) Ancestralidade Leste-africana (E-AFR); (c) Ancestralidade Oeste-africana (W-AFR). LWK: Luhya do Quênia; ESN: Esan da Nigéria; YRI: Yoruba da Nigéria; MSL: Mende de Serra Leoa; GWD: Mandika de Gâmbia; ACB: Afro-Caribenhos de Barbados; ASW: Indivíduos com ancestralidade africana no sudeste dos EUA. Fonte: Do Autor, 2024.

Populações nativo-americanas, africanas, sul-asiáticas e miscigenadas são sub-representadas em estudos genômicos e até mesmo, mais especificamente, nos estudos farmacogenéticos de câncer de mama (SCUDELER et al., 2023; SIRUGO; WILLIAMS; TISHKOFF, 2019). Com tais observações levadas em conta, pode-se constatar que é fundamental que estudos futuros busquem avaliar o potencial papel dessas variantes na clínica e no contexto terapêutico, especialmente considerando uma maior inclusão de uma diversidade étnica e populacional, visto que possivelmente o papel e as implicações clínicas será bastante específico a diferentes populações.

Além disso, destacamos que os resultados aqui encontrados podem ser de grande utilidade na prática clínica. Idealmente espera-se ter disponíveis os dados genéticos específicos do paciente no contexto da implementação farmacogenética e medicina personalizada; no entanto, na ausência deste dado, informações de ancestralidade podem

ser utilizados como *proxy* para guiar as decisões clínicas sobre o paciente e seu tratamento, uma vez que as análises de regressão linear permitem a predição das chances de um indivíduo ser carreador de determinadas variantes em acordo com sua análise de ancestralidade. Este é um fato já estabelecido por Ramos e colaboradores (2014) em um estudo em que os autores incluíram 1478 indivíduos de 19 populações de diferentes partes do mundo (1000 Genomes e dados originais de mais populações africanas) e analisaram 212 genes envolvidos na absorção, distribuição, metabolismo e excreção (ADME) de fármacos (RAMOS et al., 2014). Resultados como estes podem ser utilizados para atenuar o acesso atualmente desigual aos avanços científicos farmacogenéticos e genômicos para populações não-europeias enquanto se trabalha para que sejam adequadamente representadas nos estudos.

8. Conclusão

A partir das análises realizadas foi possível observar diferenças subcontinentais relevantes na inferência da ancestralidade genômica de 2654 indivíduos de 26 populações do Projeto *1000 Genomes* e 7 populações de Nativo-americanos peruanos e sua associação com a frequência de variantes do *ESR1*, sendo este o primeiro estudo a avaliar a sequência do gene em populações nativas. Destaca-se que as populações africanas, europeias, asiáticas e nativas foram separadas subcontinentalmente. No total, foram encontradas 145 associações entre a frequência das variantes e os 9 componentes de ancestralidade, sendo 106 associações positivas e mais de 50% das associações com os componentes africanos (W-AFR e E-AFR). Nossos resultados ecoam as recentes constatações de que a população europeia deve passar a ser considerada miscigenada para evitar interpretações errôneas em estudos genéticos. Além disso, destacamos que a especificidade das associações aos componentes subcontinentais ressalta e enfatiza a importância de se considerar as diferenças subcontinentais no design de painéis farmacogenéticos e sua implementação clínica, além de indicar possíveis implicações clínicas específicas a diferentes etnias. No contexto da prática clínica, destacamos que os resultados de associação entre variantes clinicamente relevantes do *ESR1* e os componentes de ancestralidade subcontinental podem ser utilizados para guiar as decisões clínicas sobre um paciente na ausência de seus dados genéticos. Adicionalmente, nossos dados reforçam os perigos de se extrapolar dados interpopulacionalmente, especialmente fazendo coro aos esforços que visam incluir populações negligenciadas nos estudos

genômicos e farmacogenômicos para que a implementação Farmacogenética possa ocorrer de maneira assertiva, democrática e com máxima eficácia.

Referências

- ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. Fast model-based estimation of ancestry in unrelated individuals. **Genome research**, v. 19, n. 9, p. 1655–64, set. 2009.
- AUTON, A. et al. A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68–74, 1 out. 2015.
- BELFIELD, K. D.; TICHY, E. M. Review and drug therapy implications of glucose-6-phosphate dehydrogenase deficiency. **American Journal of Health-System Pharmacy**, v. 75, n. 3, p. 97–104, 1 fev. 2018.
- BERGSTRÖM, A. et al. Insights into human genetic variation and population history from 929 diverse genomes. **Science (New York, N.Y.)**, v. 367, n. 6484, 20 mar. 2020.
- BORGES-OSÓRIO, M. R.; ROBINSON, W. M. **Genética Humana**. 3ª ed ed. Porto Alegre: [s.n.].
- BORRIE, A. E. et al. Genetic and clinical predictors of arthralgia during letrozole or anastrozole therapy in breast cancer patients. **Breast Cancer Research and Treatment**, v. 183, n. 2, p. 365–372, 6 set. 2020.
- BROCKMÖLLER, J.; TZVETKOV, M. V. Pharmacogenetics: data, concepts and tools to improve drug discovery and drug treatment. **European journal of clinical pharmacology**, v. 64, n. 2, p. 133–57, 26 fev. 2008.
- BUSTAMANTE, C. D.; DE LA VEGA, F. M.; BURCHARD, E. G. Genomics for the world. **Nature**, v. 475, n. 7355, p. 163–165, 13 jul. 2011.
- CAULFIELD, T. et al. Race and ancestry in biomedical research: exploring the challenges. **Genome Medicine**, v. 1, n. 1, p. 8, 21 jan. 2009.
- CHAN, C. W. H. et al. Pharmacogenomics of breast cancer: highlighting CYP2D6 and tamoxifen. **Journal of Cancer Research and Clinical Oncology**, v. 146, n. 6, p. 1395–1404, 8 jun. 2020.
- DALY, A. K. Pharmacogenetics: a general review on progress to date. **British Medical Bulletin**, v. 124, n. 1, p. 1–15, 11 out. 2017.
- DANECEK, P. et al. The variant call format and VCFtools. **Bioinformatics**, v. 27, n. 15, p. 2156–2158, 1 ago. 2011.
- ENGEL, R. et al. Ethnic Differences in Pharmacogenetically Relevant Genes. **Current Drug Targets**, v. 7, n. 12, p. 1641–1648, 1 dez. 2006.
- GOUVEIA, M. H. et al. Origins, Admixture Dynamics, and Homogenization of the African Gene Pool in the Americas. **Molecular Biology and Evolution**, v. 37, n. 6, p. 1647–1656, 1 jun. 2020.
- GOUVEIA, M. H. et al. Unappreciated subcontinental admixture in Europeans and European Americans and implications for genetic epidemiology studies. **Nature Communications**, v. 14, n. 1, p. 6802, 7 nov. 2023.
- GRINSHUPUN, A. et al. ESR1 activating mutations: From structure to clinical application. **Biochimica et biophysica acta. Reviews on cancer**, v. 1878, n. 1, p. 188830, 1 jan. 2023.
- GRINSHUPUN, A.; SANDUSKY, Z. M.; JESELSON, R. The Clinical Utility of ESR1 Mutations in Hormone Receptor-Positive, HER2-Negative Advanced Breast Cancer. **Hematology/Oncology Clinics of North America**, v. 37, n. 1, p. 169–181, 1 fev. 2023.

HAMAD, E.; FELDMAN, A. M. Pharmacogenetics in Heart Failure: How It Will Shape the Future. **Heart Failure Clinics**, v. 6, n. 1, p. 1–10, jan. 2010.

HARRIS, D. N. et al. Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. **Proceedings of the National Academy of Sciences**, v. 115, n. 28, 10 jul. 2018.

HENRY, N. L. et al. Genetic associations with toxicity-related discontinuation of aromatase inhibitor therapy for breast cancer. **Breast Cancer Research and Treatment**, v. 138, n. 3, p. 807–816, 2 abr. 2013.

HOMBURGER, J. R. et al. Genomic Insights into the Ancestry and Demographic History of South America. **PLoS genetics**, v. 11, n. 12, p. e1005602, dez. 2015.

JOMBART, T.; AHMED, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. **Bioinformatics**, v. 27, n. 21, p. 3070–3071, 1 nov. 2011.

KEHDY, F. S. G. et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. **Proceedings of the National Academy of Sciences**, v. 112, n. 28, p. 1–6, 14 jul. 2015.

KOZYRA, M.; INGELMAN-SUNDBERG, M.; LAUSCHKE, V. M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. **Genetics in Medicine**, v. 19, n. 1, p. 20–29, 1 jan. 2017.

LAM, Y. W. F. Principles of Pharmacogenomics. Em: **Pharmacogenomics**. [s.l.] Elsevier, 2019. p. 1–53.

LAZARIDIS, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. **Nature**, v. 513, n. 7518, p. 409–413, 17 set. 2014.

LIANG, J.; SHANG, Y. Estrogen and Cancer. **Annual Review of Physiology**, v. 75, n. 1, p. 225–240, 10 fev. 2013.

MA, Q.; LU, A. Y. H. Pharmacogenetics, Pharmacogenomics, and Individualized Medicine. **Pharmacological Reviews**, v. 63, n. 2, p. 437–459, jun. 2011.

MARIANA R. BOTTON et al. Pharmacogenomics in Latin American Populations. Em: **PHARMACOGENOMICS: Challenges and Opportunities in Therapeutic Implementation**. 2. ed. [s.l.] Elsevier, 2019. p. 330–336.

MCGUIRE, A. L. et al. The road ahead in genetics and genomics. **Nature Reviews Genetics**, v. 21, n. 10, p. 581–596, 24 out. 2020.

MITRA-GHOSH, T. et al. PharmGKB summary: lamotrigine pathway, pharmacokinetics and pharmacodynamics. **Pharmacogenetics and genomics**, v. 30, n. 4, p. 81–90, 1 jun. 2020.

MOREIRA, R. G. et al. Population genetics of PDE4B (phosphodiesterase-4B) in neglected Native Americans: Implications for cancer pharmacogenetics. **Clinical and Translational Science**, 28 mar. 2022.

NATIONAL LIBRARY OF MEDICINE. **ESR1 estrogen receptor 1 [Homo sapiens (human)] - Gene - NCBI**. Disponível em: <<https://www.ncbi.nlm.nih.gov/gene/2099#gene-expression>>. Acesso em: 18 nov. 2023.

NEWMAN, L. A. Breast Cancer Disparities: high-risk breast cancer and African ancestry. **Surgical Oncology Clinics of North America**, v. 23, n. 3, p. 579–592, jul. 2014.

NTUKIDEM, N. et al. Estrogen Receptor Genotypes, Menopausal Status, and the Lipid Effects of Tamoxifen. **Clinical Pharmacology & Therapeutics**, v. 83, n. 5, p. 702–710, 22 maio 2008.

OESTERREICH, S. et al. Associations between genetic variants and the effect of letrozole and exemestane on bone mass and bone turnover. **Breast Cancer Research and Treatment**, v. 154, n. 2, p. 263–273, 4 nov. 2015.

ONITILLO, A. A. et al. Estrogen receptor genotype is associated with risk of venous thromboembolism during tamoxifen therapy. **Breast Cancer Research and Treatment**, v. 115, n. 3, p. 643–650, 12 jun. 2009.

OZDEMIR, V.; SHEAR, N. H.; KALOW, W. What Will Be the Role of Pharmacogenetics in Evaluating Drug Safety and Minimising Adverse Effects? **Drug Safety**, v. 24, n. 2, p. 75–85, 2001.

PADHUKASAHASRAM, B. Inferring ancestry from population genomic data and its applications. **Frontiers in Genetics**, v. 5, 3 jul. 2014.

PAGANI, L. et al. Genomic analyses inform on migration events during the peopling of Eurasia. **Nature**, v. 538, n. 7624, p. 238–242, 13 out. 2016.

PENA, S. D. J. et al. The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. **PLoS ONE**, v. 6, n. 2, p. 1–9, 16 fev. 2011.

PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, set. 2007.

RAMOS, E. et al. Pharmacogenomics, ancestry and clinical decision making for global populations. **The Pharmacogenomics Journal**, v. 14, n. 3, p. 217–222, 9 jun. 2014.

RODEIRO, I. et al. Pharmacogenetics in Latin American populations: regulatory aspects, application to herbal medicine, cardiovascular and psychiatric disorders. **Drug Metabolism and Drug Interactions**, v. 27, n. 1, p. 57–60, 1 jan. 2012.

RODRIGUES-SOARES, F. et al. Genomic Ancestry, CYP2D6, CYP2C9, and CYP2C19 Among Latin Americans. **Clinical pharmacology and therapeutics**, v. 107, n. 1, p. 257–268, 1 jan. 2020.

SAYERS, E. W. et al. Database resources of the national center for biotechnology information. **Nucleic acids research**, v. 50, n. D1, p. D20–D26, 7 jan. 2022.

SCHÄRFE, C. P. I. et al. Genetic variation in human drug-related genes. **Genome medicine**, v. 9, n. 1, p. 117, 22 dez. 2017.

SCUDELER, M. M. et al. Breast cancer pharmacogenetics: a systematic review. **Pharmacogenomics**, v. 24, n. 2, p. 107–122, jan. 2023.

SCUDELER, M. M.; RODRIGUES-SOARES, F. FARMACOGENÉTICA NA AMÉRICA LATINA. **Acta Biologica Brasiliensia**, v. 3, n. 2, p. 85–100, 2020.

SIRUGO, G.; WILLIAMS, S. M.; TISHKOFF, S. A. The Missing Diversity in Human Genetic Studies. **Cell**, v. 177, n. 4, p. 1080, 2 maio 2019.

- SOLLIS, E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. **Nucleic Acids Research**, v. 51, n. D1, p. D977–D985, 6 jan. 2023.
- SOSA-MACÍAS, M. et al. Pharmacogenetics and ethnicity: relevance for clinical implementation, clinical trials, pharmacovigilance and drug regulation in Latin America. **Pharmacogenomics**, v. 17, n. 16, p. 1741–1747, nov. 2016.
- SUAREZ-KURTZ, G. et al. Self-reported skin color, genomic ancestry and the distribution of GST polymorphisms. **Pharmacogenetics and Genomics**, v. 17, n. 9, p. 765–771, set. 2007.
- SUAREZ-KURTZ, G. Pharmacogenetic testing in oncology: a Brazilian perspective. **Clinics**, v. 73, n. Suppl 1, p. 1–12, 9 out. 2018.
- SUAREZ-KURTZ, G.; PARRA, E. J. Population Diversity in Pharmacogenetics: A Latin American Perspective. Em: **Advances in Pharmacology**. [s.l.] Academic Press Inc., 2018. v. 83p. 133–154.
- SUN, Y.-S. et al. Risk Factors and Preventions of Breast Cancer. **International journal of biological sciences**, v. 13, n. 11, p. 1387–1397, 2017.
- TAN-KOI, W. C. et al. The Importance of Ethnicity Definitions and Pharmacogenomics in Ethnobridging and Pharmacovigilance. Em: **Pharmacogenomics**. [s.l.] Elsevier, 2019. p. 289–327.
- TISHKOFF, S. A. et al. The genetic structure and history of Africans and African Americans. **Science**, v. 324, n. 5930, p. 1035–1044, 22 maio 2009.
- TSUKADA, H. Another Origins Debate: Settler Colonialism in Early Virginia History. **SocArXiv Papers**, p. 1–10, 5 dez. 2021.
- URBAN, T. J. Race, Ethnicity, Ancestry, and Pharmacogenetics. **Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine**, v. 77, n. 2, p. 133–139, mar. 2010.
- WANG, C.-C. et al. Genomic insights into the formation of human populations in East Asia. **Nature**, v. 591, n. 7850, p. 413–419, 18 mar. 2021.
- WANG, J. et al. Indications of Clinical and Genetic Predictors for Aromatase Inhibitors Related Musculoskeletal Adverse Events in Chinese Han Women with Breast Cancer. **PLoS ONE**, v. 8, n. 7, p. e68798, 19 jul. 2013.
- WANG, K.; LI, M.; HAKONARSON, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. **Nucleic Acids Research**, v. 38, n. 16, p. e164–e164, 1 set. 2010.
- WANG, Y. et al. Genetic structure, divergence and admixture of Han Chinese, Japanese and Korean populations. **Hereditas**, v. 155, n. 1, p. 19, 6 abr. 2018.
- WHIRL-CARRILLO, M. et al. Pharmacogenomics Knowledge for Personalized Medicine. **Clinical Pharmacology & Therapeutics**, v. 92, n. 4, p. 414–417, out. 2012.
- WHIRL-CARRILLO, M. et al. An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. **Clinical Pharmacology & Therapeutics**, v. 110, n. 3, p. 563–572, 22 set. 2021.

WILL, M. et al. Therapeutic resistance to anti-oestrogen therapy in breast cancer. **Nature Reviews Cancer**, v. 23, n. 10, p. 673–685, 27 out. 2023.

WOJCIK, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. **Nature**, v. 570, n. 7762, p. 514–518, 19 jun. 2019.

WORLD HEALTH ORGANIZATION. **Global Breast Cancer Initiative Implementation Framework Assessing, strengthening and scaling up services for the early detection and management of breast cancer**. Geneva: [s.n.].

YANG, H.-C. et al. Genetic ancestry plays a central role in population pharmacogenomics. **Communications Biology**, v. 4, n. 1, p. 171, 5 dez. 2021.

ZHANG, F.; FINKELSTEIN, J. Inconsistency in race and ethnic classification in pharmacogenetics studies and its potential clinical implications. **Pharmacogenomics and personalized medicine**, v. 12, p. 107–123, jul. 2019.

APÊNDICE A. Tabela com valores de p (significativos - $<0,01$), regressão linear e correlação de Pierson entre as variantes analisadas e os nove componentes de ancestralidade.

SNP	Ancestralidade	Beta	p_regr	R_sqr_ajustado	p_corr	R_corr
rs9340973.G	AFRO	0,073423286	3,70233E-16	0,937368591	3,70233E-16	0,969470911
rs33979243.T	AFRO	0,039034227	3,82617E-14	0,90792488	3,82617E-14	0,95478159
rs9340773.A	SAS	0,083854804	8,91101E-14	0,901230729	8,91101E-14	0,951410269
rs9340772.T	AFRO	0,159256364	8,69678E-13	0,880681084	8,69678E-13	0,940985569
rs9340974.C	AFRL	0,165782993	1,71894E-12	0,873746669	1,71894E-12	0,937441626
rs79415092.T	EASO	0,01376695	1,18615E-10	0,820734046	1,18615E-10	0,909892677
rs9340971.G	AFRO	0,36357044	1,9199E-10	0,813453802	1,9199E-10	0,906043956
rs148633135.C	AFRO	0,047327533	3,65181E-10	0,803275448	3,65181E-10	0,90063557
rs746432.C	EURN	0,252359313	5,02901E-10	0,708776725	5,02901E-10	0,847276491
rs35151682.G	EURN	0,065477113	8,95072E-10	0,788160104	8,95072E-10	0,892543388
rs3757323.G	EURS	0,629860165	1,8035E-09	0,684083251	1,8035E-09	0,833040005
rs867240.G	AFRO	0,320775185	3,63063E-09	0,762226209	3,63063E-09	0,87848572
rs116356962.T	AFRL	0,048195183	5,82999E-09	0,752764621	5,82999E-09	0,873300656
rs867240.G	AFRL	0,200779178	6,72754E-09	0,749831578	6,72754E-09	0,871687051
rs2295190.T	EURN	0,34906033	8,65273E-09	0,650938042	8,65273E-09	0,813539322
rs2207396.A	NAT	-0,327875112	1,34985E-08	0,640934416	1,34985E-08	-0,807561277
rs9340972.A	AFRL	0,147551999	2,247E-08	0,723732193	2,247E-08	0,857194789
rs9340846.A	AFRL	0,049623523	4,17247E-08	0,614275315	4,17247E-08	0,791409636
rs35630684.T	AFRL	0,027850226	4,39836E-08	0,708055444	4,39836E-08	0,848370925
rs185717042.G	AFRL	0,016520837	6,7367E-08	0,697655031	6,7367E-08	0,842465922
rs55740371.AT	EURS	0,459326293	7,64613E-08	0,694496605	7,64613E-08	0,840664464
rs2234693.C	NAT	-0,349779463	1,09E-07	0,590119393	1,08843E-07	-0,776484489
rs150547091.A	SAS	0,008247939	1,131E-07	0,684526458	1,131E-07	0,834952334
rs2273207.G	EASO	0,427094603	1,4971E-07	0,581767298	1,4971E-07	0,771256812
rs74426520.T	AFRL	0,022791697	4,38625E-07	0,647493134	4,38625E-07	0,813383925
rs6557169.C	AFRL	0,594661674	4,43401E-07	0,647180852	4,43401E-07	0,813199618
rs9340804.T	AFRL	0,054168905	4,60687E-07	0,646075764	4,60687E-07	0,812547066
rs9397459.A	AFRO	0,29048501	6,61159E-07	0,635469053	6,61159E-07	0,806256963
rs9340844.A	EASO	0,119985872	6,77834E-07	0,53991493	6,77834E-07	0,744508287
rs111825996. AGGAG	EURS	-0,445946559	9,98684E-07	0,622982747	9,98684E-07	-0,798788731
rs1801132.C	AFRL	0,54256983	1,4458E-06	0,517432966	1,4458E-06	0,72973501
rs4870061.C	AFRL	0,526504072	1,7982E-06	0,604459609	1,7982E-06	0,787579345
rs9340971.G	AFRL	0,198630354	2,07777E-06	0,599775289	2,07777E-06	0,784719235
rs9341068.G	AFRL	0,046187635	3,1701E-06	0,585771872	3,1701E-06	0,776106306
rs73780865.G	AFRL	0,104299203	3,64727E-06	0,581020762	3,64727E-06	0,773162293
rs2813543.G	AFRL	0,232345056	3,75471E-06	0,580030566	3,75471E-06	0,772547308
rs13203975.A	NAT	-0,313140654	3,97178E-06	0,48580422	3,97178E-06	-0,70842984
rs6914438.T	EURS	-0,525222202	4,13196E-06	0,484527531	4,13196E-06	-0,70755639
rs114141096.C	AFRL	0,024151146	4,26268E-06	0,575676141	4,26268E-06	0,769837058

rs71660056.CT	AFRL	0,363319072	4,37693E-06	0,574762984	4,37693E-06	0,769267486
rs2250122.T	AFRL	-0,380336741	4,61718E-06	0,480926371	4,61718E-06	-0,705086819
rs2234693.C	AFRL	0,415205309	5,40274E-06	0,47578957	5,40274E-06	0,701549104
rs9340974.C	AFRO	0,212886501	6,85633E-06	0,558977929	6,85633E-06	0,759354207
rs9340972.A	AFRO	0,206873662	7,24315E-06	0,557009629	7,24315E-06	0,758108992
rs9341017.G	EURS	0,038045558	9,66066E-06	0,456370529	9,66066E-06	0,688010864
rs9322354.G	AFRL	0,542597493	1,16462E-05	0,449982477	1,16462E-05	0,683498738
rs974276.G	AFRL	0,542597493	1,16462E-05	0,449982477	1,16462E-05	0,683498738
rs9479119.T	EURS	-0,244989063	1,16591E-05	0,539582013	1,16591E-05	-0,746993127
rs2273206.T	AFRL	0,54228902	1,17773E-05	0,449597713	1,17773E-05	0,683226013
rs9341066.A	NAT2	0,232530939	1,1957E-05	0,449076848	1,1957E-05	0,682856644
rs185732707.C	AFRL	0,013277266	1,30345E-05	0,535406165	1,30345E-05	0,74430499
rs78574022.A	AFRL	0,013277266	1,30345E-05	0,535406165	1,30345E-05	0,74430499
rs3757323.C	EURN	0,852294189	1,41867E-05	0,443162617	1,41867E-05	0,678648499
rs9340844.A	EASL	0,149233973	1,48198E-05	0,441642972	1,48198E-05	0,677563007
rs9340848.A	AFRL	0,082457901	1,80184E-05	0,523075948	1,80184E-05	0,736310336
rs9397459.A	AFRL	0,166925583	1,9373E-05	0,520273447	1,9373E-05	0,734481116
rs74426520.T	AFRO	0,032401328	2,36106E-05	0,512545803	2,36106E-05	0,729413443
rs2250122.T	AFRO	-0,568938514	2,44148E-05	0,423990744	2,44148E-05	-0,664824062
rs6914438.T	EURN	-0,814737388	2,84474E-05	0,418482703	2,84474E-05	-0,660798849
rs2273206.T	NAT	-0,396907918	3,27924E-05	0,413317717	3,27924E-05	-0,657001932
rs148633135.C	AFRL	0,023885386	3,30091E-05	0,49918692	3,30091E-05	0,720568833
rs9322354.G	NAT	-0,396598446	3,35606E-05	0,412472291	3,35606E-05	-0,656378345
rs974276.G	NAT	-0,396598446	3,35606E-05	0,412472291	3,35606E-05	-0,656378345
rs6557169.C	AFRO	0,834143503	3,42827E-05	0,497656139	3,42827E-05	0,719548395
rs2228480.A	NAT	0,176169018	4,83363E-05	0,399003589	4,83363E-05	0,64636269
rs9341016.C	SAS	0,191170045	5,40325E-05	0,478913995	5,40325E-05	0,706935241
rs139960913.T	EURN	0,008515431	6,13779E-05	0,473546734	6,13779E-05	0,703281497
rs1801132.C	AFRO	0,753207196	6,3689E-05	0,388633721	6,3689E-05	0,638544374
rs9340772.T	AFRL	0,074823629	6,67137E-05	0,470008796	6,67137E-05	0,700862643
rs111825996. AGGAG	EURN	-0,646089891	7,13988E-05	0,467111927	7,13988E-05	-0,698875848
rs9322354.G	AFRO	0,794309999	8,3388E-05	0,378345176	8,3388E-05	0,630691596
rs974276.G	AFRO	0,794309999	8,3388E-05	0,378345176	8,3388E-05	0,630691596
rs2273206.T	AFRO	0,793860709	8,40959E-05	0,37801993	8,40959E-05	0,630441756
rs2813543.G	EURS	-0,22087306	9,97649E-05	0,452617223	9,97649E-05	-0,688848702
rs4870061.C	AFRO	0,721819074	0,000128055	0,441564217	0,000128055	0,681103258
rs2295191.T	EASL	0,17854054	0,000165658	0,429950113	0,000165658	0,672868567
rs2234693.C	AFRO	0,571958039	0,000168382	0,350779296	0,000168382	0,609153053
rs746432.C	EURS	0,109144299	0,000171022	0,350156775	0,000171022	0,608657848
rs71660056.C	AFRL	-0,264730134	0,000171514	0,428366162	0,000171514	-0,671737683
rs2273207.G	NAT	-0,227983503	0,000185077	0,346988906	0,000185077	-0,606131588
rs71660056.C	EASL	0,391074843	0,000195748	0,422303557	0,000195748	0,667391501
rs3756933.C	AFRL	-0,097439154	0,000200878	0,343688382	0,000200878	-0,603488293
rs33979243.T	AFRL	0,017133363	0,000214329	0,41810991	0,000214329	0,664368507

rs9340799.G	NAT	-0,189957393	0,000241559	0,336204878	0,000241559	-0,597451651
rs139548761.G	AFRL	0,005239546	0,000247249	0,411446166	0,000247249	0,659536443
rs34788032.G	EURS	0,008133755	0,000265316	0,408131525	0,000265316	0,657119673
rs144000393.G	SAS	0,090689855	0,000293341	0,403383116	0,000293341	0,653641945
rs9341016.C	AFRL	-0,184564984	0,000300431	0,402248829	0,000300431	-0,652808452
rs117509702.G	EURN	0,047475847	0,000318147	0,399519326	0,000318147	0,650798396
rs139548761.G	AFRO	0,008188806	0,000323447	0,398730222	0,000323447	0,650216128
rs55740371.AT	EURN	0,587664861	0,000329932	0,397780764	0,000329932	0,649514845
rs2813558.A	AFRL	0,202964351	0,000408308	0,314495573	0,000408308	0,579583977
rs12681.T	AFRL	0,202964351	0,000408308	0,314495573	0,000408308	0,579583977
rs2813543.G	AFRO	0,305835302	0,000412957	0,386952246	0,000412957	0,641462513
rs35630684.T	AFRO	0,033150751	0,000466338	0,381015226	0,000466338	0,637004409
rs2207396.A	SAS	0,298275413	0,000472941	0,308309411	0,000472941	0,574390757
rs2295191.T	EASO	0,124031756	0,000490909	0,378492198	0,000490909	0,635100394
rs13203975.A	EASO	0,369999026	0,000556308	0,301419568	0,000556308	0,568550971
rs139740651.C	AFRL	0,005713297	0,000568486	0,371232752	0,000568486	0,6295899
rs71660056.C	EASO	0,27084268	0,000592891	0,369139294	0,000592891	0,627991817
rs9340805.G	EURN	0,02947133	0,000657187	0,363985503	0,000657187	0,624040129
rs9341068.G	AFRO	0,058680833	0,000692859	0,361325242	0,000692859	0,62199054
rs9340799.G	SAS	0,22778703	0,000699145	0,29162281	0,000699145	0,560142479
rs201145204.T	AFRO	0,010549395	0,000762127	0,356505082	0,000762127	0,618259556
rs75311867. CAA	AFRO	0,374208024	0,000850542	0,350913444	0,000850542	0,613903011
rs9397459.A	EURS	-0,148489848	0,000873515	0,349549373	0,000873515	-0,612835539
rs9340846.A	AFRO	0,05356816	0,001225448	0,267080995	0,001225448	0,538502288
rs2250122.T	NAT	0,217566237	0,001515455	0,257614104	0,001515455	0,529918544
rs111825996. AGGAG	AFRO	0,489473758	0,001524295	0,320485394	0,001524295	0,589632071
rs71660056.C	AFRO	-0,368082373	0,00154099	0,319905935	0,00154099	-0,589160163
rs185732707.C	AFRO	0,016637942	0,00157011	0,318909114	0,00157011	0,588347473
rs78574022.A	AFRO	0,016637942	0,00157011	0,318909114	0,00157011	0,588347473
rs2273207.G	EASL	0,399077872	0,001614296	0,254779387	0,001614296	0,527321089
rs9322335.C	EURS	0,30778204	0,001648357	0,253840739	0,001648357	0,526458181
rs200880341.C	EASL	0,004829152	0,001673644	0,315499684	0,001673644	0,585559302
rs867239.G	AFRO	0,363300859	0,001726117	0,313846325	0,001726117	0,584202424
rs3756933.C	AFRO	-0,134232081	0,001747464	0,251211071	0,001747464	-0,524033134
rs4870061.C	EASO	-0,426993735	0,001747768	0,313177802	0,001747768	-0,583652884
rs9341016.C	AFRO	-0,260624985	0,001835137	0,310556422	0,001835137	-0,581493048
rs75311867.CAA	EASO	-0,243744359	0,001929602	0,307850336	0,001929602	-0,579254972
rs114141096.C	AFRO	0,028802555	0,001934534	0,307712467	0,001934534	0,579140715
rs200075329.C	NAT	0,025002354	0,002169523	0,241404395	0,002169523	0,514888831
rs2207396.A	NAT2	-0,398109046	0,002620746	0,232759069	0,002620746	-0,506690585
rs6557169.C	EASO	-0,452138955	0,002635524	0,290845967	0,002635524	-0,564988609
rs2813558.A	SAS	-0,171139451	0,00291321	0,22788621	0,00291321	-0,502010723
rs12681.T	SAS	-0,171139451	0,00291321	0,22788621	0,00291321	-0,502010723
rs71660056.CT	AFRO	0,417095624	0,003114429	0,281601124	0,003114429	0,55707906

rs9340973.G	AFRL	0,026518818	0,003245934	0,279296342	0,003245934	0,555089622
rs867239.G	EURN	-0,381777727	0,003511837	0,27489229	0,003511837	-0,551268173
rs2813558.A	AFRO	0,273735015	0,003579113	0,218339223	0,003579113	0,492713022
rs12681.T	AFRO	0,273735015	0,003579113	0,218339223	0,003579113	0,492713022
rs71660056.CT	EASO	-0,284354595	0,003592236	0,273622267	0,003592236	-0,550161228
rs150886628.A	AFRL	0,002247595	0,003740384	0,271350428	0,003740384	0,54817553
rs1801132.C	EASO	-0,404115727	0,004224729	0,210586475	0,004224729	-0,485031595
rs6914438.T	NAT	0,249184355	0,005732544	0,196178966	0,005732544	0,470423611
rs4986934.C	EURN	-0,071613314	0,006530072	0,189976535	0,006530072	-0,463993285
rs142712646.T	EURS	0,004255691	0,006954241	0,235801429	0,006954241	0,516109844
rs111825996. AGGAG	AFRL	0,269248581	0,007205632	0,233727628	0,007205632	0,514177521
rs4870061.C	EASL	-0,506166439	0,007884876	0,228448857	0,007884876	-0,509225788
rs2813558.A	EASO	-0,177760929	0,008115201	0,179560289	0,008115201	-0,452988996
rs12681.T	EASO	-0,177760929	0,008115201	0,179560289	0,008115201	-0,452988996
rs13203975.A	AFRO	0,417039677	0,008169472	0,179239503	0,008169472	0,452645853
rs9322335.C	EASL	-0,378035534	0,008539319	0,177106633	0,008539319	-0,450357692
rs867239.G	EURS	-0,210345406	0,008821573	0,221835107	0,008821573	-0,502952982
rs9341066.A	SAS	-0,100383252	0,009278673	0,173097648	0,009278673	-0,446025051