



UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE
TECNOLOGIA



Programa de Pós-Graduação em Biotecnologia

**ANÁLISE DE SNPS RELACIONADOS COM DOENÇAS
CARDIOVASCULARES EM INDIVÍDUOS COM
ANCESTRALIDADE AFRICANA E EUROPEIA**

Hugo Santos Lintz

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biotecnologia do Programa de Pós-Graduação em Biotecnologia da Universidade Federal de São Carlos.

Orientador:

Prof. Dr. Euclides Matheucci Junior

Bolsista: CAPES – 88887.753197/2022-00

**São Carlos - SP
2024**

HUGO SANTOS LINTZ

**ANÁLISE DE SNPS RELACIONADOS COM DOENÇAS
CARDIOVASCULARES EM INDIVÍDUOS COM
ANCESTRALIDADE AFRICANA E EUROPEIA**

Dissertação apresentada como parte dos requisitos para
obtenção do título de Mestre em Biotecnologia do
Programa de Pós-Graduação em Biotecnologia da
Universidade Federal de São Carlos.

Orientador:

Prof. Dr. Euclides Matheucci Junior

BANCA EXAMINADORA

Prof. Dr. Euclides Matheucci Junior
UFSCar

Dr. Calógeras Antônio de Albergaria Barbosa
Professor Adjunto IV aposentado - UFSCar

Dra. Adriana Medaglia
DNAConsult

São Carlos
2024

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Biotecnologia

Folha de aprovação

Assinatura dos membros da comissão examinadora, que avaliou e aprovou a Defesa de Mestrado do candidato Hugo Santos Lintz realizada em 30/08/2024:

Prof. Dr. Euclides Matheucci Junior
Universidade Federal de São Carlos (UFSCar)

Dr. Calógeras Antonio de Albergaria Barbosa
Professor Adjunto IV Aposentado (UFSCar)

Dra. Adriana Medaglia
DNA Consult Genética e Biotecnologia LTDA (DNAConsult)

AGRADECIMENTOS

Agradeço a todos aqueles que me apoiaram durante meu trajeto nos últimos anos.

Ao meu orientador, Prof. Dr. Euclides Matheucci Junior, que sempre me deu auxílio para me aperfeiçoar como pesquisador.

Às pessoas da empresa DNA Consult Genética Biotecnologia Ltda. que me ajudaram na realização desta pesquisa.

Aos meus pais, que sempre me deram apoio incondicional em todas as decisões que tomei.

Aos meus amigos, que estão sempre presentes no meu dia a dia.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.753197/2022-00.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 88887.753197/2022-00.

Lintz, Hugo. **Análise de SNPs relacionados a doenças cardiovasculares em indivíduos com ancestralidade africana e europeia**. 2024. Dissertação (Programa de Pós-graduação em Biotecnologia) – UFSCar, São Carlos, 2024.

RESUMO

Doenças cardiovasculares (CVD) são os tipos de doenças não transmissíveis mais comuns no mundo, responsáveis por mais de 17 milhões de mortes no ano de 2017, das quais mais de três quartos ocorreram em países de baixa e média renda, além de serem complexas e heterogêneas, envolvendo a interação de fatores genéticos, epigenéticos e ambientais. No que tange às CVD, 5,3% da população adulta no Brasil recebeu diagnóstico médico de alguma condição cardíaca totalizando mais de 8 milhões de pessoas. A espécie do ser humano moderno, caracterizada como *Homo sapiens sapiens*, foi originada de populações localizadas na África subsaariana há 200 mil anos. Mesmo tendo suas origens vindas de uma mesma região, os anos de isolamento geográfico permitiram uma diferenciação genética entre populações. Abordagens baseadas em ancestralidade genética utilizam marcadores genéticos com grandes diferenças de frequência alélica entre grupos ancestrais, chamados marcadores informativos de ancestralidade. Os mais amplamente usados são baseados em Polimorfismos de Nucleotídeo Único (SNP), que demonstram capacidade superior em prever origem biogeográfica. A identificação de SNPs associados ao risco para o desenvolvimento de doenças mais complexas é um objetivo importante dentro dos estudos genéticos modernos. Cada variante genética é valiosa para indicar um gene ou via de relevância biológica para o distúrbio. Dessa forma, o presente estudo teve como objetivo analisar SNPs relacionados a doenças cardiovasculares em indivíduos com ancestralidades africana e europeia na população brasileira. Os resultados demonstram que características como hipertensão e cardiomiopatias estão mais associadas à ancestralidade africana e europeia, respectivamente, enquanto outras apresentam uma distribuição equilibrada entre as duas populações.

Palavras-Chave: Ancestralidade; SNP; CVD.

Lintz, Hugo. **Análise de SNPs relacionados a doenças cardiovasculares em indivíduos com ancestralidade africana e europeia.** 2024. Dissertação (Programa de Pós-graduação em Biotecnologia) – UFSCar, São Carlos, 2024.

ABSTRACT

Cardiovascular diseases (CVD) are the most common type of non-communicable disease in the world, responsible for more than 17 million deaths in 2017, of which more than three-quarters occurred in low and middle-income countries, in addition to being complex and heterogeneous, involving the interaction of genetic, epigenetic and environmental factors. Regarding CVD, 5,3% of the adult population in Brazil has received a medical diagnosis of some cardiac condition, totaling more than 8 million people. The modern human species, characterized as *Homo sapiens sapiens*, originated from populations located in sub-Saharan Africa 200 thousand years ago. Even though their origins came from the same region, years of geographic isolation allowed genetic differentiation between populations. Approaches based on genetic ancestry use genetic markers with large differences in allele frequency between ancestral groups, called ancestry-informative markers. The most widely used are based on Single Nucleotide Polymorphisms (SNPs), which demonstrate superior ability to predict biogeographic origin. The identification of SNPs associated with the risk for the development of more complex diseases is an important objective within modern genetic studies. Each genetic variant is valuable to indicate a gene or pathway of biological relevance to the disorder. Thus, the present study aims to analyze SNPs related to cardiovascular diseases in individuals with African and European ancestries in the Brazilian population. The results show that traits such as hypertension and cardiomyopathies are more associated with African and European ancestry, respectively, while others are more evenly distributed between the two populations.

Keywords: Ancestry; SNP; CVD.

LISTA DE FIGURAS

CAPÍTULO 1 - INTRODUÇÃO

Figura 1. Migrações humanas ao longo do tempo	15
Figura 2. Ocupação no território brasileiro ao longo dos anos (séculos XVI-XX)	16

LISTA DE GRÁFICOS

CAPÍTULO 1 - INTRODUÇÃO

Gráfico 1. Diagnóstico médico autorreferido de alguma doença do coração (percentual 2013 - 2019) 13

LISTA DE TABELAS

CAPÍTULO 3 – MATERIAIS E MÉTODOS

Tabela 1. Lista de SNPs relacionados com doenças cardiovasculares	20
---	----

CAPÍTULO 4 – RESULTADOS

Tabela 2. Frequência Alélica SNPs Doença Coronariana	26
Tabela 3. Frequência Alélica SNPs Remodelamento Vascular	26
Tabela 4. Frequência Alélica SNPs Colesterol	27
Tabela 5. Frequência Alélica SNPs Sinalização e Transporte Celular	28
Tabela 6. Frequência Alélica SNPs Pressão Sanguínea	29
Tabela 7. Frequência Alélica SNPs Aortopatias	30
Tabela 8. Frequência Alélica Insuficiência Cardíaca	30
Tabela 9. Frequência Alélica Cardiomiopatias	31

LISTA DE ABREVIATURAS E SIGLAS

AIMS	<i>Ancestry Informative Markers</i>
CVD	<i>Cardiovascular Diseases</i>
DNA	<i>Deoxyribonucleic acid (ácido desoxirribonucleico)</i>
GWAS	<i>Genome-wide Association Studies</i>
SNP	<i>Single Nucleotide Polimorphism</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Doenças Cardiovasculares	12
1.2	Ancestralidade	14
1.3	SNP	17
2	OBJETIVOS	18
2.1	Objetivo Geral	18
2.2	Objetivos Específicos	18
3	MATERIAIS E MÉTODOS	19
3.1	Seleção dos SNPs	19
3.2	Alcance do Banco de Marcadores	22
3.3	Determinação de Etnia	23
3.4	Análise de Frequência Alélica	24
4	RESULTADOS	25
5	DISCUSSÃO	31
6	CONCLUSÃO	344
	REFERÊNCIAS	35

1 INTRODUÇÃO

1.1 Doenças Cardiovasculares

Doenças cardiovasculares (CVD) são os tipos de doenças não transmissíveis mais comuns no mundo, sendo responsáveis por mais de 17 milhões de mortes no ano de 2017, das quais mais de três quartos ocorreram em países de baixa e média renda (KAPTOGE et al., 2019).

Segundo dados da *American Heart Association*, e da Organização Mundial da Saúde, CVD são a causa mais frequente de morte prematura, com incidência e prevalência aumentando a cada ano. Fatores de risco tradicionais, como aumento da pressão sistólica, colesterol HDL baixo, diabetes e tabagismo, são usados para calcular o risco de eventos cardiovasculares em 10 anos com escalas especialmente projetadas (HOJMAN; KARSULOVIC, 2022).

A hipertensão é considerada um fator de risco para mortes prematuras em todo o mundo, representando 17,9 milhões de mortes em 2018. Está presente em aproximadamente 22% da população global (afetando 1 em 4 homens e 1 em 5 mulheres), com maior prevalência observada na África (27%) e menor prevalência na América, onde se tem uma média de 18% na América do Norte (< 20%) e América do Sul (20 – 24,9%) (SINGH et al., 2021).

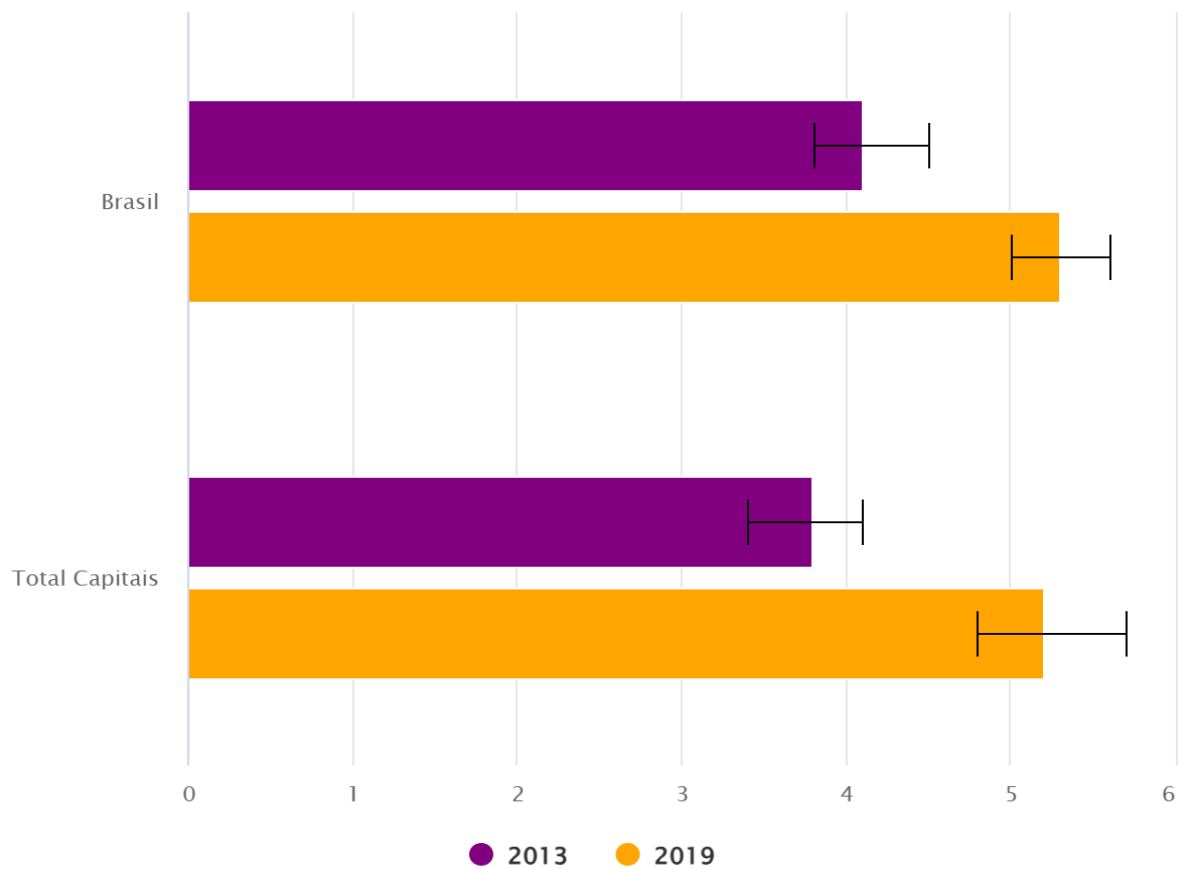
Acidente Vascular Cerebral (AVC) é a segunda principal causa de morte e incapacidade em todo mundo, com mais de 13 milhões de novos casos e 5,5 milhões de mortes registradas em 2016. Embora a incidência geral em países desenvolvidos tenha diminuído 42% no período de 1970 a 2008, ela aumentou mais de 100% nos países de baixa renda e em desenvolvimento, onde há um aumento desproporcional entre os indivíduos com ancestralidade africana (PRAPIADOU; DEMEL; HYACINTH, 2018).

Doenças cardiovasculares podem ser influenciadas por diversos fatores, alguns mais diretos como fatores genéticos familiares e hábitos de saúde, e outros que podem atuar de forma indireta como, por exemplo, um desregulamento no remodelamento de vasos sanguíneos e alterações no índice de colesterol, que podem aumentar significativamente as chances de um caso de hipertensão e desenvolvimento de AVC (PRAPIADOU; DEMEL; HYACINTH, 2018).

Ao se tratar da realidade brasileira, a Pesquisa Nacional de Saúde (PNS), conduzida pelo Ministério da Saúde em convênio com o Instituto Brasileiro de Geografia e Estatística (IBGE), tem como um de seus principais objetivos a investigação das condições de saúde da população brasileira. Até o momento foram realizadas duas edições da pesquisa, nos anos de 2013 e 2019.

Conforme dados da PNS, aproximadamente 14,6% dos brasileiros com 18 ou mais anos foram diagnosticados com colesterol alto, o que equivale a cerca de 23,2 milhões de indivíduos. No estado de São Paulo, esse percentual é ainda maior, atingindo 16,3%. No que tange às CVDs, 5,3% da população adulta no Brasil recebeu diagnóstico médico de alguma condição cardíaca totalizando mais de 8 milhões de pessoas. Em São Paulo, essa prevalência sobe para 6,1% (FUNDAÇÃO OSWALDO CRUZ, 2021).

Gráfico 1. Diagnóstico médico autorreferido de alguma doença do coração (percentual 2013 - 2019)



Fiocruz | ICICT | LIS | PCDaS | IBGE

Legenda: Percentual de diagnóstico médico autorreferido de alguma doença do coração nos anos de 2013 e 2019 no Brasil e nas capitais brasileiras. O gráfico mostra que no país no ano de 2013 cerca de 4,10% da população autodeclarou possuir alguma doença do coração e em 2019 esse número sobe para 5,30%. Nas capitais brasileiras em 2013 o percentual foi de 3,80% e em 2019 de 5,20%.

Fonte: Fundação Oswaldo Cruz, 2021.

Existem inúmeros fatores ambientais e comportamentais que estão envolvidos no desenvolvimento e progressão de CVDs. Além disso, muitas variações genéticas têm sido associadas a esta enfermidade. Para algumas variações, o caminho já está mais bem elucidado, para muitas outras, especialmente aquelas que envolvem múltiplas células ou sistemas de órgãos, o papel da genética é muito mais complexo (WIMLEY, 2017).

Estudos genéticos populacionais em grande escala fornecem uma oportunidade para melhorar nossa compreensão da herdabilidade de características complexas. Estudos de associação genômica ampla (GWAS), por exemplo, já foram capazes de identificar 208 *loci* de suscetibilidade para Doença Arterial Coronariana (CAD). Esses *loci* explicam uma fração modesta (~15%) da hereditariedade, e foram amplamente identificados em populações europeias. Além disso, todos são exclusivamente autossômicos (TCHEANDJIEU et al., 2022).

Estudos do poder preditivo das principais variantes genéticas relacionadas às CVDs, individualmente ou em combinação, mostram melhorias. A maior parte da hereditariedade inexplicável de um indivíduo pode estar escondida entre os milhares de SNPs que não atingiram significância no genoma. De fato, avanços recentes mostram que modelos de predição genômica que consideram todas as variantes genéticas disponíveis podem estratificar com mais eficiência aqueles com maior risco (ABRAHAM et al., 2016).

1.2 Ancestralidade

A espécie do ser humano moderno (*Homo sapiens sapiens*) originou na África Subsaariana há 200 mil anos e sua migração, para fora no continente africano, se deu a partir de 70 mil anos atrás, como pode ser observado na Figura 1. Milhares de anos de isolamento geográfico foram acompanhados de diversificação genética, dando origem aos diversos grupos populacionais humanos (NORRIS et al., 2018).

Populações de diferentes origens biogeográficas diferem geneticamente umas das outras devido às variantes que são raras em alguns grupos. As frequências dos alelos variantes podem ser medidas para comparar a distância genética de populações, além de revelar a variação dentro de um mesmo grupo populacional (YAHYA et al., 2020).

Estudos anteriores mostraram que variantes comuns a qualquer população (Menor frequência alélica ou $MAF > 5\%$) são normalmente compartilhadas entre todas as populações, embora a frequência com que ocorrem possa variar substancialmente entre diferentes grupos ancestrais. Variantes de frequência mais baixa ($MAF < 5\%$) são mais prováveis de serem específicas de uma população e são mais prováveis de ocorrer em populações ancestrais (FISH et al., 2018).

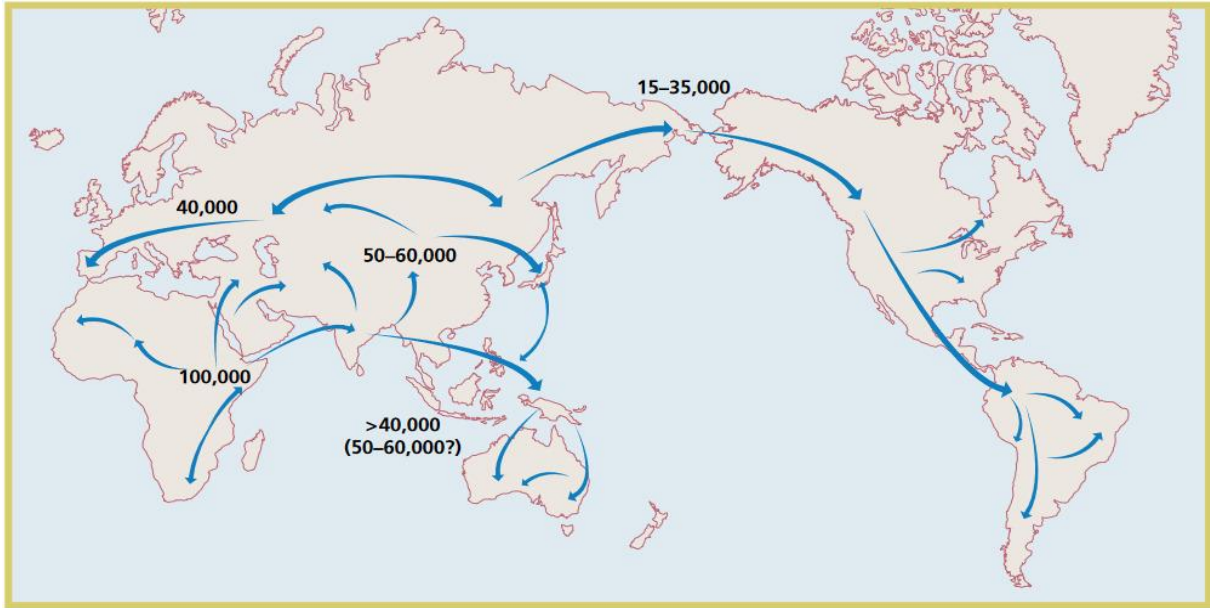


Figura 1. Migrações humanas ao longo do tempo

Legenda: Rotas de migração do ser humano moderno para fora do continente africano. Cerca de 100 mil anos atrás para todo o continente africano, entre 100 mil e 60 mil para os continentes da Europa e Ásia, entre 60 mil e 40 mil anos para a Oceania e, por fim, entre 40 mil e 15 mil anos atrás para as américas. Fonte: Cavalli-Sforza e Feldman (2003).

A população brasileira tem uma das constituições genéticas mais heterogêneas de todas, com uma extensa miscigenação, sendo que os principais grupos populacionais podem ser divididos entre aqueles com maior descendência africana, e os com maior descendência europeia (SOUZA et al., 2019).

Ademais, os anos de exploração colonial realizados no Brasil permitiu que o território brasileiro fosse influenciado de diversas formas, conforme mostrado na Figura 2, onde denota migrações exploratórias ao longo dos séculos de história do povo brasileiro (SOUZA et al., 2019)

Compreender como a diversidade genética humana é distribuída e suas implicações tem sido foco em estudos de genética clínica, populacional e forense. Entende-se que a maior parte da variação genética ocorre entre indivíduos do mesmo grupo populacional, enquanto a variação entre indivíduos de populações distantes é restrita a uma porção do genoma (ESCHER et al., 2022).

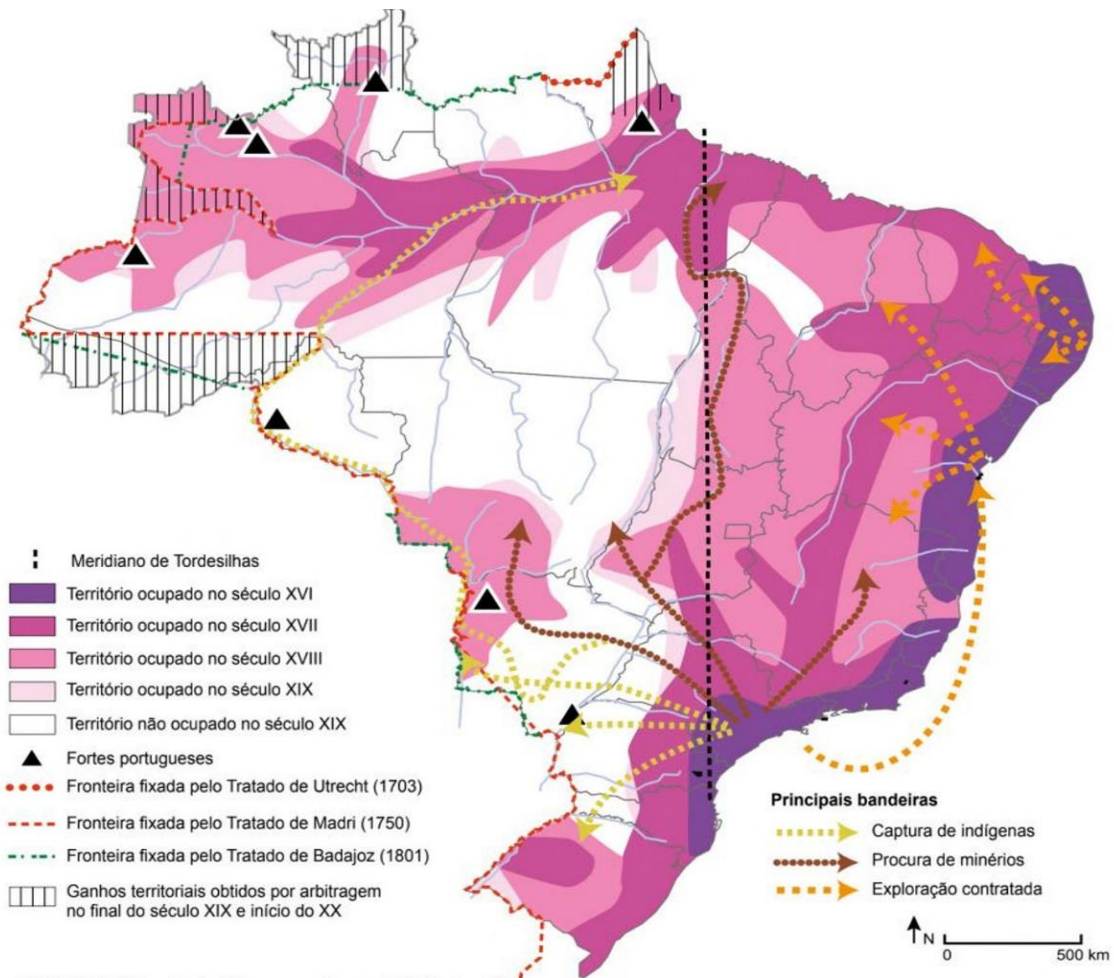


Figura 2. Ocupação no território brasileiro ao longo dos anos (séculos XVI-XX)

Legenda: Ocupação territorial entre os séculos XV e XX no território brasileiro diferenciando entre rotas conquistadas de indígenas, rotas de procura de minérios e por exploração contratada. Fonte: Théry e Mello-Théry (2023).

Inferir a ancestralidade biogeográfica de forma precisa tem importância para várias áreas de aplicação. Identificação de marcadores informativos de ancestralidade (AIMs) no genoma é essencial para detectar a estratificação da população em estudos de associação de controle de caso para doenças complexas como câncer, diabetes, doenças neurodegenerativas e doenças cardiovasculares (TOMA; DAWSON; ADJEROH, 2019).

Com o desenvolvimento de pesquisas sobre a diversidade da genética humana, mais e mais variações específicas em populações têm sido discernidas. Geralmente marcadores genéticos com diferença em frequência alélica são selecionados como AIMs para distinguir populações diferentes (JIN et al., 2020).

Avanços recentes baseados no genoma agora oferecem a oportunidade de medir a variabilidade individual nos genes, ambientes e estilo de vida que constituem uma fase importante para a medicina de precisão. As abordagens de sequenciamento de próxima geração também nos permite procurar padrões ou variantes genéticas, expressão gênica e redes de

regulação em subpopulações que se correlacionam com genótipos de doença (MENSAH et al., 2020).

1.3 SNP

A genética é uma ferramenta essencial para desvendar a complexidade dos processos biológicos. Isso levou a investimentos significativos e desenvolvimentos tecnológicos em genética para aplicações acadêmicas e industriais. Seu avanço levou a descobertas significativas de pesquisa e aplicações clínicas, melhorando nossa compreensão de doenças e características. Isso também auxiliou o tratamento e o desenvolvimento de medicamentos, abrindo caminho para a medicina personalizada com base na genética (KOCKUM; HUANG; STRIDH, 2023).

Genotipagem é o processo de determinar a sequência genética de um indivíduo, ou seja, identificar as variações no DNA de uma pessoa ou organismo. Essa técnica é usada para analisar o genoma e identificar diferenças em *loci* específicos de genes, o que pode fornecer informações sobre características herdadas, predisposição a doenças, resposta a medicamentos, entre outros fatores biológicos (KOCKUM; HUANG; STRIDH, 2023).

As frequências alélicas variantes podem ser medidas para comparar a distância genética das populações e revelar a variação genética dentro e entre as populações. Tais alelos podem ser escolhidos como candidatos para inferir a ancestralidade de um indivíduo particular, denominados de AIMs (YAHYA et al., 2020).

Estes, por sua vez, podem ser encontrados entre todos os tipos de polimorfismos de DNA, como sequências curtas repetidas em tandem (STRs), polimorfismos de inserção-deleção (INDELs) ou polimorfismos de nucleotídeo único (SNPs) (YAHYA et al., 2020).

SNP é considerada a forma mais simples e, também, a mais comum de variação do DNA. Este tipo de alteração consiste na substituição de um único nucleotídeo por outro e podem influenciar a atividade promotora (expressão gênica), conformação do RNA mensageiro e, também, a eficiência translacional (SHASTRY, 2009).

SNPs autossômicos possuem algumas características como: ampla distribuição no genoma, variações bialélicas comuns e diferenças de frequência distintas entre regiões geográficas, que os tornam marcadores ideais para inferir a ancestralidade de indivíduos desconhecidos (JIN et al., 2020).

Estudos de SNPs são bastante comuns em genotipagem, e há diversas técnicas que podem ser usadas para identificar e analisar essas variações. Algumas das principais técnicas

incluem: microarranjos de SNP (*SNP Arrays*); PCR em tempo real (qPCR) específica para SNPs e *TaqMan Assay* (KOCKUM; HUANG; STRIDH, 2023).

Os recentes avanços em pesquisas de SNPs tornaram possíveis conduzir estudos de ampla associação genômica (GWAS) como primeiro passo, não apenas para identificar mutações casuais, mas os SNPs usados se tornam úteis para criar preditores de doenças permitindo a criação de um “perfil genômico” (TOMA; DAWSON; ADJEROH, 2019).

A maioria das características humanas são poligênicas e seus fenótipos são tipicamente influenciados por numerosos genes e fatores ambientais e, possivelmente, suas interações. Essas características foram denominadas como complexas, que se distinguem das mendelianas, que são moldadas por um ou poucos genes principais (YU et al., 2020).

Estudos genéticos populacionais em grande escala fornecem uma oportunidade para melhorar nossa compreensão da base herdada de características complexas. Um dos fatores que complica a precisão de um diagnóstico clínico, por exemplo, é a presença de múltiplos genes que têm um pequeno efeito positivo no desenvolvimento de uma enfermidade (SINGH et al., 2021).

Os GWAS descobriram com sucesso milhares de associações entre SNPs e características complexas, que revolucionaram a compreensão da arquitetura poligênica delas. Posteriormente, a fim de aumentar o poder e a precisão na identificação de mais variantes, houve numerosos estudos de acompanhamento usando meta-análises estatísticas de múltiplos GWAS, combinando diversas fontes de dados e abrangendo diferentes populações (YU et al., 2021).

2 OBJETIVOS

2.1 Objetivo Geral

- Analisar as diferenças de frequências alélicas de marcadores de doenças cardiovasculares entre indivíduos populações brasileiras de diferentes ancestralidades.

2.2 Objetivos Específicos

- Selecionar 2 grupos de SNPs, um relacionado às doenças cardiovasculares e outro à ancestralidade;

- Separar dois grupos de indivíduos em ancestralidade europeia e africana pela análise de genomas;
- Determinar as frequências alélicas de doenças cardiovasculares dos dois grupos;
- Analisar as diferenças das frequências alélicas obtidas de ambos os grupos para cada característica de interesse.

3 MATERIAIS E MÉTODOS

3.1 Seleção dos SNPs

O processo de seleção de variantes genéticas se iniciou pela análise de bancos de dados públicos, o objetivo da etapa era identificar características que estão relacionadas com doenças cardiovasculares. Cada variante estudada nestes bancos de dados tem uma avaliação de seu impacto clínico diante de uma determinada característica.

A análise mostrou as variantes que possuíam associação a alguma CVD ou característica relacionada a elas, e esta associação possui algum tipo de risco patogênico dentro do genoma.

Após a análise foi feita uma revisão bibliográfica para encontrar GWAS que relacionavam as variantes selecionadas com as características desejadas. Os GWAS escolhidos deveriam fornecer informações de *odds ratio*, *p-value* e serem implementados em populações que tinham identificadores de ancestralidade europeia e africana.

O *odds ratio* (OR) é uma medida estatística usada para quantificar a força e a direção da associação entre duas variáveis binárias, neste caso analisa-se a relação entre uma determinada variante genética e a presença ou ausência de uma característica específica. Um $OR > 1$ sugere uma associação positiva indicando que a presença da variante está associada a maiores chances de se apresentar a característica observada (PARK; HAN, 2022).

O OR é calculado pela razão entre as chances de a característica aparecer no grupo com a variante dividido pela chance da característica aparecer no grupo sem a variante.

O valor *p* é uma medida estatística que indica a probabilidade de se obter os resultados observados, ou resultados mais extremos, se a hipótese nula (que não há associação) for verdadeira. Este dado funciona como um índice de compatibilidade entre os dados e o modelo, que varia entre 0 (rejeitando hipótese nula) a 1 (acetando hipótese nula). Quanto mais o valor *p* se aproxima de 0, mais ele sugere que os resultados são improváveis de ocorrer ao acaso; fornecendo evidências para a existência de uma correlação entre a ocorrência da característica e a presença da variante (MANSOURNIA; NAZEMIPOUR; ETMINAN, 2022).

Para realizar a filtragem de SNPs, foi utilizado o *software* PLINK v1.9, o qual realiza a filtragem baseada no desvio do equilíbrio de Hardy-Weinberg, *call rate* de genotipagem e menor frequência alélica (CHANG et al., 2015) As diretrizes de associação dos SNPs foram seguidas conforme mostrado em (NIKPAY et al., 2015)

Foram selecionados 90 SNPs que seguiam estas diretrizes e que analisavam diversas características de doenças cardiovasculares, sendo elas: doença coronariana, remodelamento vascular, colesterol LDL, sinalização e transporte celular, pressão sanguínea, aortopatias, insuficiência cardíaca e cardiomiopatias.

As variantes relacionadas a estas características foram estudadas nos seguintes GWAS selecionados para este estudo: HARST; VERWEIJ, 2018; KESSLER; VILNE; SCHUNKERT, 2016; NIKPAY et al., 2015; PEDEN et al., 2011; RHODES et al., 2019; SCHUNKERT et al., 2011; SEIFERT et al., 2019; TEMPRANO-SAGRERA et al., 2022; TESLOVICH et al., 2010.

Tabela 1. Lista de SNPs relacionados com doenças cardiovasculares

rsID	Gene	Cromossomo	Alelo risco/ref	OR	P-value
rs17087335	NOA1	chr4:56972417	T/G	1,6	4×10^{-19}
rs17114036	PLPP3	chr1:56497149	G/A	1,07	5×10^{-8}
rs10455872	LPA	chr6:160589086	G/A	1,07	2×10^{-12}
rs2306374	MRAS	chr3:138401110	C/T	1,6	3×10^{-8}
rs7692387	GUCY1A1	chr4:155714157	G/A	1,07	3×10^{-23}
rs6544713	ABCG8	chr2:43846742	T/C	1,6	2×10^{-18}
rs602633	PSRC1	chr1:109278889	G/T	1,07	4×10^{-58}
rs599839	PSRC1	chr1:109279544	G/A	1,06	3×10^{-10}
rs515135	TDRD15	chr2:21063185	A/G	1,07	6×10^{-17}
rs11556924	ZC3HC1	chr7:130023656	T/C	1,6	1×10^{-23}
rs2252641	TEX41	chr2:145043894	G/A	1,06	5×10^{-13}
rs2505083	JCAD	chr10:30046193	C/T	1,07	4×10^{-8}
rs9319428	FLT1	chr13:28399484	A/G	1,06	5×10^{-11}
rs4773144	COL4A2	chr13:110308365	G/A	1,07	4×10^{-9}
rs2895811	HHIPL1	chr14:99667605	C/T	1,07	1×10^{-14}
rs2023938	HDAC9	chr7:18997152	C/T	1,8	4×10^{-11}
rs17609940	ANKS1A	chr6:35067023	G/C	1,07	1×10^{-8}
rs3825807	ADAMTS7	chr15:78796769	A/G	1,08	1×10^{-12}
rs3798220	LPA	chr6:160540105	C/T	1,51	3×10^{-11}
rs116843064	ANGPTL4	chr19:8364439	G/A	1,14	8×10^{-10}
rs2075650	TOMM40	chr19:44892362	G/A	1,14	2×10^{-14}
rs17114036	PLPP3	chr1:56497149	A/G	1,17	4×10^{-19}

rs11206510	PCSK9	chr1:55030366	T/C	1,08	$4x10^{-8}$
rs579459	Y_RNA	chr9:133278724	C/T	1,1	$3x10^{-10}$
rs264	LPL	chr8:19955669	G/A	1,11	$6x10^{-12}$
rs964184	ZPR1	chr11:116778201	G/C	1,13	$8x10^{-10}$
rs2954029	TRIB1AL	chr8:125478730	T/A	1,06	$3x10^{-29}$
rs6544713	ABCG8	chr2:43846742	T/C	1,06	$2x10^{-20}$
rs515135	TDRD15	chr2:21063185	C/T	1,07	$5x10^{-29}$
rs599839	PSRC1	chr1:109279544	A/T	1,11	$2x10^{-19}$
rs1122608	SMARCA4	chr19:11052925	G/T	1,14	$1x10^{-9}$
rs17696736	NAA25	chr12:112049014	G/A	1,218	$1x10^{-34}$
rs216172	SMG6	chr17:2223210	C/G	1,07	$5x10^{-7}$
rs11830157	KSR2	chr12:117827636	G/T	1,12	$2x10^{-9}$
rs974819	DYNC2H1	chr11:103789839	T/C	1,07	$6x10^{-29}$
rs7692387	GUCY1A1	chr4:155714157	G/A	1,08	$4x10^{-22}$
rs12936587	PEMT	chr17:17640408	A/G	1,07	$3x10^{-10}$
rs273909	MIR3936HG	chr5:132331660	G/C	1,7	$3x10^{-10}$
rs1561198	VAMP5	chr2:85582866	T/C	1,06	$2x10^{-23}$
rs10947789	KCNK5	chr6:39207146	T/C	1,7	$3x10^{-12}$
rs10953541	BCAP29	chr7:107604100	C/T	1,08	$3x10^{-8}$
rs2681492	ATP2B1	chr12:89619312	T/C	1,71	$9x10^{-15}$
rs9815354	ULK4	chr3:41871159	A/G	1,12	$3x10^{-9}$
rs2856830	HLA-DPA1	chr6:33073957	C/T	1,56	$8x10^{-20}$
rs6599230	SCN5A	chr3:38633221	C/T	1,71	$3x10^{-8}$
rs13376333	KCNN3	chr1:154841877	T/C	1,5	$2x10^{-21}$
rs3825214	TBX5	chr12:114357638	A/G	1,56	$3x10^{-13}$
rs3918226	NOS3	chr7:150993088	T/C	1,14	$1x10^{-20}$
rs381815	PLEKHA7	chr11:16880721	T/C	1,36	$2x10^{-9}$
rs2681472	ATP2B1	chr12:89615182	A/G	1,74	$1x10^{-20}$
rs2384550	TBX3-AS1	chr12:114914926	G/A	1,07	$4x10^{-8}$
rs12413409	CNNM2	chr10:102959339	G/A	1,12	$1x10^{-9}$
rs13266183	RN7SL250P	chr8:54355052	C/T	1,36	$2x10^{-12}$
rs1805124	SCN5A	chr3:38603929	C/T	3,14	$4x10^{-19}$
rs7193343	ZFH3	chr16:72995261	C/T	1,74	$1x10^{-10}$
rs2106261	ZFH3	chr16:73017721	T/C	1,33	$3x10^{-30}$
rs56062135	SMAD3	chr15:67163292	C/T	1,07	$4x10^{-17}$
rs11014166	CACNB2	chr10:18419869	A/T	1,5	$6x10^{-8}$
rs7164883	HCN4	chr15:73359833	G/A	1,21	$3x10^{-17}$
rs11047543	LOC105369698	chr12:24635405	A/G	4,7	$2x10^{-9}$
rs3184504	SH2B3	chr12:111446804	T/C	1,12	$3x10^{-34}$

rs6495122	ULK3	chr15:74833304	A/C	1,08	$2x10^{-10}$
rs17514846	FURIN	chr15:90873320	A/C	1,07	$1x10^{-26}$
rs11222084	ZBTB44-DT	chr11:130403335	T/C	0,785	$3x10^{-25}$
rs12740374	CELSR2	chr1:109274968	G/A	0,767	$2x10^{-57}$
rs7528419	CELSR2	chr1:109274570	A/C	0,777	$1x10^{-9}$
rs3729856	GATA4	chr8:11757066	G/A	1,24	$1x10^{-9}$
rs646776	CELSR2	chr1:109275908	G/A	0,77	$8x10^{-11}$
rs4936099	ADAMTS8	chr11:130410830	G/A	0,789	$5x10^{-9}$
rs629301	CELSR2	chr1:109275684	G/A	0,77	$1x10^{-170}$
rs660240	CELSR2	chr1:109275216	C/T	1,06	$2x10^{-28}$
rs870142	STX18-AS1	chr4:4646320	T/C	1,456	$3x10^{-10}$
rs4766578	ATXN2	chr12:111466567	T/A	1,04	$6x10^{-20}$
rs41306688	ADPRHL1	chr13:113424243	C/A	1,82	$4x10^{-9}$
rs233716	RPH3A	chr12:112602139	G/A	1,263	$5x10^{-13}$
rs7982677	GPC5	chr13:92336070	A/C	1,311	$3x10^{-9}$
rs9982601	LINC00310	chr21:34226827	T/C	1,18	$4x10^{-10}$
rs2306374	MRAS	chr3:138401110	C/T	1,12	$3x10^{-8}$
rs111245230	SVEP1	chr9:110407495	C/T	1,14	$8x10^{-9}$
rs1412444	LIPA	chr10:89243170	T/C	1,09	$2x10^{-21}$
rs4845625	IL6R	chr1:154449591	T/C	1,06	$6x10^{-16}$
rs4252120	PLG	chr6:160722576	T/C	1,07	$3x10^{-8}$
rs11065987	ATXN2-AS	chr12:111634620	G/A	1,266	$2x10^{-9}$
rs17465637	MIA3	chr1:222650187	C/A	1,14	$8x10^{-31}$
rs4977574	CDKN2B-AS1	chr9:22098575	G/A	1,29	$9x10^{-223}$
rs11556924	ZC3HC1	chr7:130023656	C/T	1,09	$8x10^{-12}$
rs1746048	LINC02881	chr10:44280376	C/T	1,9	$3x10^{-10}$
rs46522	UBE2Z	chr17:48911235	T/C	1,06	$2x10^{-8}$

Legenda: A tabela mostra dados dos SNPs selecionados, incluindo identificadores da variante, gene onde a variante foi encontrada, localização do cromossomo (chr) da variante, alelos de risco e referencial, além de valores de *odds ratio* e *p-value* calculados pelos GWAS selecionados previamente.

3.2 Alcance do Banco de Marcadores

Para a realização da próxima etapa, é necessária a utilização da plataforma advinda da empresa Illumina, conhecida por desenvolver e comercializar uma ampla gama de produtos voltados para a pesquisa genômica, incluindo plataformas de sequenciamento de nova geração (NGS) e microarranjos de DNA.

O *InfiniumTM Global Screening Array-24 v3.0 BeadChip* é uma plataforma desenvolvida pela empresa Illumina para a análise de genotipagem de larga escala. O BeadChip

contém milhares de marcadores genéticos, distribuídos pelo genoma humano, projetado para análises em estudos populacionais, saúde pública, genômica translacional e pesquisas clínicas. A versão 3.0 inclui variantes atualizadas para cobrir regiões genômicas de interesse, incluindo regiões de importância para diferentes etnias e áreas associadas a doença e características específicas (BACKBONE, 2016).

A plataforma é amplamente utilizada em estudos de associação do genoma (GWAS) devido à sua alta densidade de marcadores e à sua capacidade de fornecer dados preciosos sobre a diversidade genética. Ela permite identificar variantes genéticas com maior precisão, contribuindo para a medicina personalizada e para o entendimento de fatores de risco genético (BACKBONE, 2016).

O chip, referido acima, representa uma matriz de genotipagem avançada que se utiliza do sistema *iScan* integrado com um *software* de análise, e o *Infinium high-throughput screening* (HTS), fornecendo ferramentas para amplificar, fragmentar, hibridizar, rotular e detectar variantes genéticas (BACKBONE, 2016).

O chip irá identificar, dentro da lista criada, aqueles SNPs que já possuem dados computados e quantificados dentro de sua própria lista. Aqueles presentes dentro do chip serão selecionados para fazer parte do presente estudo. Ao todo foram selecionados 90 SNPs relacionados com doenças cardiovasculares e características relevantes para o desenvolvimento de CVDs.

Juntamente com os SNPs selecionados, o banco de dados contará com dados brutos de genotipagem de 1.256 indivíduos brasileiros, predominantemente do estado de São Paulo. Estes dados contêm informações de variantes presentes no genoma do indivíduo com seus dados dos alelos apresentados por elas.

O banco de dados brutos de genotipagem foi adquirido através da empresa DNA Consult Genética e Biotecnologia LTDA (DNAConsult) e será utilizado posteriormente para o cálculo de etnia separado para cada indivíduo.

3.3 Determinação de Etnia

Esta etapa foi realizada com o intuito de separar os indivíduos do banco de dados coletado em dois grupos: um grupo aqueles que possuísem maior porcentagem de ancestralidade africana, e outro com maior porcentagem de ancestralidade europeia. O processo foi realizado através da ferramenta do instituto Sanger, nomeada de HapMap.

Este projeto é uma iniciativa internacional que visa desenvolver um mapa abrangente das variações genéticas humanas, particularmente os haplótipos, que são combinações específicas de variantes genéticas (SNPs) que tendem a ser herdadas em conjunto.

Este projeto determina a etnia através da análise de SNPs comparada com 11 populações de referência que são utilizadas como modelos comparativos para os dados de genoma dos indivíduos de interesse. Os dados das 11 populações de referências são: *African ancestry in Southwest USA (ASW)*; *Utah residents with Northern and Western European ancestry from the CEPH collection (CEU)*; *Han Chinese in Beijing, China (CHB)*; *Chinese in Metropolitan Denver, Colorado (CHD)*; *Gujarati Indians in Houston, Texas (GIH)*; *Japanese in Tokyo, Japan (JPT)*; *Luhya in Webuye, Kenya (LWK)*; *Mexican ancestry in Los Angeles, California (MXL)*; *Maasai in Kinyawa, Kenya (MKK)*; *Toscani in Italia (TSI)* e *Yoruba in Ibadan, Nigeria (YRI)*.

Para o cálculo de etnia são necessários dados de genotipagem de populações de referência, fornecido pelo programa, e dados da população de interesse, fornecido pela DNAConsult. Dados de variantes das populações de referência devem seguir a certos critérios determinados pelo HapMap, como mostrado em Sanger Institute (2013).

O *software* possui modelos computacionais e algoritmos próprios para o cálculo de ancestralidade. Dessa forma é necessário introduzir apenas os dados de genomas a serem analisados para que o próprio programa funcione.

Dos 1.256 dados de genomas de indivíduos presentes, 741 foram considerados como tendo maior porcentagem de ancestralidade africana, enquanto 515 foram considerados como tendo maior porcentagem de ancestralidade europeia

3.4 Análise de Frequência Alélica

A frequência alélica refere-se à frequência com que um alelo específico ocorre em uma determinada população. Um alelo é uma forma alternativa de um gene que pode ocupar um determinado locus em um cromossomo. Geralmente são expressos como proporções ou porcentagens.

Para a análise de frequência alélica, foi utilizado um *software* desenvolvido internamente, de forma integrada aos estudos e ao desenvolvimento do projeto. Esse *software* foi projetado especificamente para realizar a varredura de arquivos brutos de genotipagem e a aferição das frequências alélicas, empregando a linguagem de programação Python.

Para que o programa seja exitoso deve-se, primeiramente, criar 2 documentos separados no formato .txt: um contendo a lista de SNPs de interesse a serem analisados, e o outro contendo dados brutos dos genomas, separados por indivíduos da população de interesse.

Em um primeiro momento, o programa fará a leitura do documento contendo a lista de SNPs e, em seguida, fará uma varredura dentro da lista de exomas dos indivíduos, um número de vezes igual à quantidade de SNPs presentes na primeira lista.

Em cada varredura, o programa identificará se o SNP de interesse está presente dentro do genoma do indivíduo, dando informações de quais são os alelos atuantes dentro dos dados de cada indivíduo.

O código do programa está disponível em <https://github.com/GoLintz/mestrado>.

O cálculo da frequência alélica foi realizado a partir dos dados de frequência genotípica obtidos através do programa. O seguinte cálculo foi utilizado:

$$p^2 + 2pq + q^2$$

onde:

- $p = \text{alelo } A$
- $q = \text{alelo } B$
- $p^2 = \text{frequência do homozigoto } AA$
- $2pq = \text{frequência do heterozigoto } AB$
- $q^2 = \text{frequência do homozigoto } BB$

Logo a frequência do alelo A pode ser calculada por $p = p^2 + 0,5(2pq)$ e a frequência do alelo B $q = q^2 + 0,5(2pq)$.

4 RESULTADOS

Para cada população, os cálculos de frequência alélica foram realizados para cada SNP individualmente. Ao final, as variantes foram agrupadas de acordo com a característica à qual estavam associadas. Os alelos de risco foram destacados para permitir a análise das diferenças entre os indivíduos com maior ancestralidade africana e aqueles com maior ancestralidade europeia.

Para análise de relação de variante foi calculado um valor de influência, representado pelo logaritmo natural do *odds ratio* (lnOR) de cada variante. Quanto maior o valor apresentado mais influência a variante tem sobre a determinação da característica a qual ela está relacionada.

Doença coronariana apresentou um total de 12 SNPs relacionados, dos quais 6 destes estavam ligados aos alelos A e G, 4 aos alelos T e C e 2 aos alelos T e G. Como pode ser observado na Tabela 2, as variantes rs11556924, rs654413, rs2306374 e rs17087335 foram as que apresentaram maior influência sobre a característica, com um valor de ln(OR) maior que 0,4, enquanto que as outras variantes obtiveram valores abaixo de 0,07.

Tabela 2. Frequência Alélica SNPs Doença Coronariana

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		G	A	G	A	
rs17114036	G/A	15%	85%	12%	88%	0,067659
rs10455872	G/A	2%	98%	7%	93%	0,067659
rs2252641	G/A	14%	87%	54%	46%	0,058269
rs7692387	G/A	95%	6%	80%	21%	0,067659
rs599839	G/A	82%	18%	22%	78%	0,058269
rs515135	A/G	54%	46%	20%	80%	0,067659
rsID	Alelo risco/ref	T	C	T	C	ln(OR)
rs11556924	T/C	3%	97%	38%	62%	0,470004
rs6544713	T/C	14%	86%	31%	69%	0,470004
rs2505083	C/T	96%	4%	57%	43%	0,067659
rs2306374	C/T	91%	9%	86%	15%	0,470004
rsID	Alelo risco/ref	T	G	T	G	ln(OR)
rs17087335	T/G	21%	79%	19%	81%	0,470004
rs602633	G/T	82%	18%	21%	79%	0,067659

Legenda: SNPs encontrados relacionados à característica de Doença Coronariana, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

Estas variantes apresentaram resultados de frequência alélica semelhantes entre as duas populações, com diferenças maiores no SNP rs11556924, cujos alelos patogênicos representaram 3% na população de ancestralidade africana e 38% na população de ancestralidade europeia.

Tabela 3. Frequência Alélica SNPs Remodelamento Vascular

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		G	A	G	A	
rs9319428	A/G	69%	31%	69%	31%	0,05827
rs3825807	A/G	11%	89%	43%	57%	0,04696
rs4773144	G/A	37%	63%	45%	55%	0,06766

rsID	Alelo risco/ref	T	C	T	C	ln(OR)
rs2895811	C/T	77%	23%	57%	43%	0,06766
rs2023938	C/T	78%	22%	90%	10%	0,58779
rsID	Alelo risco/ref	G	C	G	C	ln(OR)
rs17609940	G/C	99%	1%	81%	19%	0,06766

Legenda: SNPs encontrados relacionados à característica de Remodelamento Vascular, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

De forma semelhante, a Tabela 3 de remodelamento vascular se distribuiu de forma equivalente entre as duas populações, onde os alelos patogênicos dos SNPs encontrados apresentaram uma frequência semelhante, não adquirindo nenhuma preferência dentro dos indivíduos.

A variante com mais influência na característica, rs2023938, obteve um ln de *odds ratio* igual a 0, 58779 com maior presença na população de ancestralidade africana, com 22% do alelo patogênico, em comparação à população de ancestralidade europeia, com apenas 10%.

Uma das características com o maior número de variantes relacionadas foi o colesterol, o que pode ser verificado na Tabela 4. Além disso, também apresentou um alto número de SNPs com alta relevância na determinação da característica.

Tabela 4. Frequência Alélica SNPs Colesterol

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		T	C	T	C	
rs3798220	C/T	100%	0%	99%	1%	0,41211
rs579459	C/T	87%	13%	79%	21%	0,09531
rs515135	C/T	54%	46%	20%	80%	0,067659
rs11206510	T/C	87%	13%	83%	17%	0,076961
rs6544713	T/C	14%	86%	31%	69%	0,058269
rsID	Alelo risco/ref	T	A	T	A	ln(OR)
rs2954029	T/A	29%	71%	45%	55%	0,058269
rs599839	A/T	82%	18%	22%	78%	0,10436
rsID	Alelo risco/ref	G	A	G	A	ln(OR)
rs116843064	G/A	100%	0%	97%	3%	0,131028
rs2075650	G/A	13%	87%	13%	87%	0,131028
rs264	G/A	89%	11%	85%	15%	0,10436
rs17696736	G/A	2%	98%	43%	57%	0,19721
rs17114036	A/G	15%	85%	12%	88%	0,157004
rsID	Alelo risco/ref	T	G	T	G	ln(OR)
rs1122608	G/T	1%	99%	25%	75%	0,131028
rsID	Alelo risco/ref	G	C	G	C	ln(OR)
rs964184	G/C	22%	78%	16%	84%	0,122218

Legenda: SNPs encontrados relacionados à característica de Colesterol, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

Por mais que algumas dessas variantes não apresentem diferenças significativas nas frequências alélicas entre as populações, é possível destacar algumas que podem ser consideradas relevantes. Por exemplo, rs599839 e rs17696736 mostram uma maior distribuição de alelos patogênicos na população de ancestralidade europeia, enquanto os SNPs rs1122608 e rs964184 apresentam uma maior frequência de alelos patogênicos na população de ancestralidade africana.

Tabela 5. Frequência Alélica SNPs Sinalização e Transporte Celular

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		G	C	G	C	
rs216172	C/G	63%	37%	63%	37%	0,067659
rs273909	G/C	1%	99%	13%	87%	0,530628
rsID	Alelo risco/ref	T	C	T	C	ln(OR)
rs1561198	T/C	72%	28%	48%	52%	0,058269
rs10947789	T/C	97%	3%	74%	26%	0,530628
rs974819	T/C	53%	47%	28%	72%	0,067659
rs10953541	C/T	1%	99%	21%	79%	0,076961
rsID	Alelo risco/ref	G	A	G	A	ln(OR)
rs7692387	G/A	95%	6%	80%	21%	0,076961
rs12936587	A/G	74%	26%	55%	45%	0,067659
rsID	Alelo risco/ref	T	G	T	G	ln(OR)
rs11830157	G/T	83%	17%	59%	41%	0,113329

Legenda: SNPs encontrados relacionados à característica de Sinalização e Transporte Celular, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

Sinalização e transporte celular apresentaram 2 SNPs com um ln do *odds ratio* acima de 0,5, significando uma relevância para a determinação da característica elevada (Tabela 5). O identificador rs10947789 teve uma distribuição de alelos patogênicos elevada nas duas populações estudadas, com maior incidência na de ancestralidade africana, enquanto o rs216172 com distribuição igual para as duas populações.

Existem muitos fatores que podem influenciar a pressão sanguínea de um indivíduo; por este motivo, esta foi a característica com o maior número de SNPs relacionados dentro dos estudos observados. Esta também foi a característica que apresentou 2 SNPs com altíssima relevância, rs1805124 e 11047543 com ln do *odds ratio* acima de 1,0 (Tabela 6).

De modo geral, a distribuição de frequência alélica apresentou-se com uma maior quantidade de alelos considerados patogênicos na população de ancestralidade africana, apesar

de a população de ancestralidade europeia ter apresentado certa quantidade de alelos patogênicos.

Tabela 6. Frequência Alélica SNPs Pressão Sanguínea

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		T	C	T	C	
rs2681492	T/C	87%	13%	85%	15%	0,536493
rs13376333	T/C	30%	70%	30%	70%	0,405465
rs3918226	T/C	0%	100%	10%	90%	0,131028
rs381815	T/C	19%	81%	30%	70%	0,307485
rs2106261	T/C	28%	72%	18%	82%	0,285179
rs3184504	T/C	2%	98%	46%	54%	0,113329
rs2856830	C/T	90%	10%	87%	13%	0,444686
rs6599230	C/T	22%	78%	21%	79%	0,536493
rs13266183	C/T	18%	82%	23%	77%	0,307485
rs1805124	C/T	69%	31%	78%	22%	1,144223
rs7193343	C/T	25%	75%	17%	83%	0,553885
rs56062135	C/T	4%	96%	22%	79%	0,067659
rsID	Alelo risco/ref	G	A	G	A	ln(OR)
rs9815354	A/G	81%	19%	82%	18%	0,113329
rs11047543	A/G	96%	4%	85%	16%	1,547563
rs2681472	A/G	9%	91%	15%	86%	0,553885
rs3825214	A/G	24%	76%	20%	80%	0,444686
rs2384550	G/A	69%	31%	63%	37%	0,067659
rs12413409	G/A	96%	4%	91%	9%	0,113329
rs7164883	G/A	33%	67%	16%	84%	0,19062
rsID	Alelo risco/ref	A	T	A	T	ln(OR)
rs11014166	A/T	87%	13%	68%	32%	0,405465
rsID	Alelo risco/ref	C	A	C	A	ln(OR)
rs6495122	A/C	22%	78%	53%	47%	0,076961
rs17514846	A/C	12%	88%	54%	46%	0,067659

Legenda: SNPs encontrados relacionados à característica de Pressão Sanguínea, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

A característica aortopatias foi a que obteve os SNPs com menor relevância nos cálculos realizados, conforme demonstrado na Tabela 7. Nesse contexto, foram identificadas 4 variantes de ln do *odds ratio* inferior a 0. Dentre elas, o SNP rs3729856 se destacou como a única variante de alta relevância, apresentando uma maior distribuição de alelos patogênicos na população de ancestralidade europeia. Os dados sugerem, portanto, uma presença mais influente de fatores externos para a determinação desta característica, quando comparado às outras referenciadas no presente estudo.

Tabela 7. Frequência Alélica SNPs Aortopatias

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		G	A	G	A	
rs3729856	G/A	1%	99%	12%	88%	0,215111
rs646776	G/A	60%	40%	79%	21%	-0,26136
rs4936099	G/A	15%	85%	36%	64%	-0,23699
rs629301	G/A	40%	60%	21%	79%	-0,26136
rs12740374	G/A	75%	25%	79%	21%	-0,26527
rsID	Alelo risco/ref	T	C	T	C	ln(OR)
rs11222084	T/C	21%	79%	33%	67%	-0,24207
rsID	Alelo risco/ref	C	A	C	A	ln(OR)
rs7528419	A/C	27%	73%	21%	79%	-0,25231

Legenda: SNPs encontrados relacionados à característica de Aortopatias, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

A característica com a menor quantidade de SNPs relacionados foi insuficiência cardíaca, com apenas 3 SNPs; isso pode ser verificado na Tabela 8. Porém, estes obtiveram uma distribuição de frequência alélica bastante divergente entre as duas populações, indicando um possível fator genético ancestral na determinação desta característica. O SNP de maior relevância foi o rs870142, onde os alelos patogênicos representaram 24% na população de ancestralidade europeia e 5% na ancestralidade africana.

Tabela 8. Frequência Alélica Insuficiência Cardíaca

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		T	C	T	C	
rs660240	C/T	39%	61%	20%	80%	0,058269
rs870142	T/C	5%	95%	24%	76%	0,375693
rsID	Alelo risco/ref	T	A	T	A	ln(OR)
rs4766578	T/A	2%	98%	48%	52%	0,039221

Legenda: SNPs encontrados relacionados à característica de Insuficiência Cardíaca, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

Por fim, cardiomiopatias se portou com uma distribuição de alelos patogênicos maior na população de ancestralidade europeia, como pode ser observado na Tabela 9. Os SNPs rs1746048, rs17465637 e rs233716 destacaram-se, todos apresentando uma porcentagem de alelos patogênicos superior a 60% na população de ancestralidade europeia.

Tabela 9. Frequência Alélica Cardiomiopatias

rsID	Alelo risco/ref	População Africana		População Europeia		ln(OR)
		T	C	T	C	
rs2306374	C/T	93%	7%	86%	15%	0,113329
rs111245230	C/T	100%	0%	97%	3%	0,131028
rs11556924	C/T	3%	97%	38%	62%	0,086178
rs1746048	C/T	54%	46%	15%	86%	0,641854
rs9982601	T/C	23%	77%	12%	88%	0,165514
rs1412444	T/C	42%	58%	36%	64%	0,086178
rs4845625	T/C	29%	71%	45%	55%	0,058269
rs4252120	T/C	85%	15%	70%	30%	0,067659
rs46522	T/C	6%	94%	56%	44%	0,058269
rsID	Alelo risco/ref	G	A	G	A	ln(OR)
rs11065987	G/A	2%	98%	42%	58%	0,235862
rs4977574	G/A	14%	86%	49%	51%	0,254642
rs233716	G/A	51%	49%	60%	40%	0,23349
rsID	Alelo risco/ref	A	C	A	C	ln(OR)
rs41306688	C/A	100%	0%	97%	3%	0,598837
rs17465637	C/A	81%	19%	26%	74%	0,131028
rs7982677	A/C	36%	64%	28%	73%	0,27079

Legenda: SNPs encontrados relacionados à característica de Cardiomiopatias, correlacionando o rsID (identificador da variante) com os alelos presentes encontrados nas análises. Em vermelho está destacado o alelo de risco para as duas populações no estudo.

O principal objetivo da análise dos dados apresentados é entender quais são as variantes mais influentes na determinação das características das quais elas estão relacionadas. Além disso, é importante entender se as distribuições das frequências alélicas, entre populações, definem ou não a existência de um fator ancestral na determinação da característica.

5 DISCUSSÃO

A população brasileira se apresenta muito diversa em sua ancestralidade genética, e essa diversidade tem repercussões significativas na predisposição às várias condições de saúde, incluindo doenças cardiovasculares. Variações genéticas únicas a grupos étnicos distintos podem influenciar desde fatores de risco até respostas de tratamento. Compreender essas diferenças genéticas é fundamental para uma abordagem mais personalizada e eficaz na prevenção e tratamento das doenças cardiovasculares (SOUZA et al., 2019).

Entre as doenças cardiovasculares, algumas demonstram uma influência genética maior, o que sugere que certas características têm uma base genética mais bem evidente. Como observado nos resultados, condições como: doença coronariana, colesterol e cardiomiopatias

demonstraram fortes vínculos com variantes genéticas específicas, facilitando a análise dessas características e sua associação com o risco cardiovascular (ABRAHAM et al., 2016; NORRIS et al., 2018; SINGH et al., 2021).

Embora muitas características mostrem alto grau de correlação genética compartilhada entre ancestrais, outras têm variantes genéticas distintas com grandes efeitos que são mais comuns em grupos ancestrais específicos, o que possibilita a identificação das variantes patogênicas exclusivas de um único grupo (ESCHER et al., 2022).

Isso indica que, embora uma característica possa ter um fator genético altamente relevante devido à grande quantidade de SNPs relacionados, isso não significa necessariamente que ela possa ser diferenciada entre populações de ancestralidades distintas, já que as frequências alélicas dessas populações tendem a se distribuir de maneira semelhante.

Foi demonstrado que variações genéticas em vários genes podem contribuir para a patogênese de cardiomiopatias, podendo também ser polimorfismos associados à doença que controlam a suscetibilidade (MAZZACCARA et al., 2018).

Embora cada polimorfismo sozinho possa determinar a doença em uma pequena porcentagem, existem aqueles que contribuem de forma mais efetiva para a determinação da característica. Neste estudo, os SNPs que apresentaram maiores resultados de associação foram os rs1805124 e rs11047543, que apresentaram um *odds ratio* de 3,14 e 4,7, respectivamente.

Os estudos de Mazzaccara et al. (2018) e Olesen et al. (2012) foram capazes de associar os SNPs citados acima como sendo de alta relevância para cardiomiopatia e pressão sanguínea. A presença do rs1805124 com alelos patogênicos resultou em um aumento de 5% no risco de desenvolvimento de cardiomiopatia dilatada, enquanto o rs11047543 foi identificado como fator de risco para infarto do miocárdio devido às alterações no gene SOX5, no qual a variante atua.

O SNP rs 11047543 está localizado próximo ao gene SOX5, conhecido por codificar um fator de transcrição expresso em vários tecidos, mais predominantemente no coração, músculo esquelético e fígado. A principal expressão do SOX5 no tecido cardíaco sugere papéis potenciais na miogênese humana. Em um estudo anterior, camundongos homocigotos para uma mutação SOX6 mostraram ultraestrutura anormal ao músculo esquelético cardíaco (SMITS et al., 2001).

Como SOX5 e SOX6 são coexpressos e interagem intimamente, foi levantada a hipótese de que SOX5 pode exercer um efeito semelhante ao de SOX6 no desenvolvimento muscular humano. Em consonância com essa hipótese, foi demonstrado que camundongos deficientes em SOX5 morrem de insuficiência cardíaca. O rs11047543 pode afetar o padrão de condução atrial

modificando as conexões musculares interatriais, devido ao papel potencial de SOX5 na miogênese (VOGEL et al., 2021).

Estudos que analisaram características relacionadas à pressão arterial trouxeram evidências que indicam que a ascendência africana, mais especificamente da África Ocidental, se apresenta como fator de risco para a hipertensão, enquanto a ascendência europeia atua como fator de proteção para a mesma característica (JORDAN et al., 2023). Além disso, esta mesma pesquisa analisou pacientes com cardiomiopatia dilatada e foi capaz de chegar à mesma conclusão que a demonstrada pelo presente trabalho.

O gene SCN5A pertence a uma família de genes que fornecem instruções para fazer canais de sódio. Esses canais abrem e fecham em momentos específicos para controlar o fluxo de átomos de sódio carregados positivamente para dentro das células. Os canais de sódio contendo proteínas produzidas pelo gene são abundantes nas células musculares cardíacas e desempenham papéis importantes na capacidade dessas células de gerar e transmitir sinais elétricos, mutações nesse gene podem alterar consideravelmente as chances de se desenvolver alguma CVD (AKAI et al., 2000)

Analisando a distribuição de frequências alélicas de doenças coronarianas demonstrada na Tabela 2, observa-se uma distribuição mais uniforme dos alelos patogênicos. Em alguns SNPs, os alelos patogênicos estão mais presentes em indivíduos com maior porcentagem de ancestralidade europeia, enquanto em outros SNPs essa prevalência é observada em indivíduos de ancestralidade africana.

Portanto, as associações genéticas podem diferir dentro de grupos de uma mesma população devido à variabilidade regional nas origens dos seus ancestrais. A estratificação genética é afetada pela proporção de ascendência, tanto europeia quanto africana, em diferentes participantes de um mesmo grupo populacional (DE OLIVEIRA et al., 2022).

Entretanto, é importante ressaltar que as doenças cardiovasculares são multifatoriais, e os fatores externos, como hábitos alimentares, estilo de vida, atividade física e histórico médico também desempenham um papel crucial no desenvolvimento e progressão dessas condições. Assim, uma abordagem integrada, considerando tanto os aspectos genéticos quanto os fatores ambientais e comportamentais é essencial para compreender plenamente a complexidade das doenças cardiovasculares e desenvolver estratégias eficazes de prevenção e tratamento personalizado (BENZEVAL; AGUIRRE; KUMARI, 2023).

6 CONCLUSÃO

Em conclusão, este estudo revela uma intrincada e diversificada relação entre a herança genética e a suscetibilidade a enfermidades cardiovasculares. As descobertas destacam a importância de considerar as influências genéticas ao analisar a prevalência e a incidência dessas doenças na população brasileira.

A história multicultural do país, com sua mistura de etnias e origens ancestrais, desempenha um papel fundamental na variabilidade genética encontrada na população. Essa diversidade pode explicar parte das diferentes taxas de incidência de CVDs em distintos grupos populacionais brasileiros.

Ademais, os avanços tecnológicos na área da genética e o desenvolvimento de ferramentas de análise genômica permitem uma melhor compreensão das contribuições genéticas para as doenças cardiovasculares. Ao identificar marcadores genéticos específicos, associados a essas condições, podemos aprimorar as estratégias de prevenção, diagnóstico e tratamento, promovendo uma medicina mais personalizada e eficaz.

As evidências sugerem que certas variantes genéticas associadas às doenças cardiovasculares podem ter uma prevalência maior em indivíduos com ancestralidade africana, enquanto outras podem ser mais comuns em indivíduos com maior ancestralidade europeia. Esta disparidade genética pode explicar, em partes, as diferenças observadas nos índices de frequência alélica entre os dois grupos analisados.

Contudo, é essencial enfatizar que o fator genético não é o único determinante dessas doenças. Fatores ambientais, comportamentais e socioeconômicos também desempenham um papel significativo na manifestação dessas enfermidades. Portanto, uma abordagem integrada, considerando tanto as influências genéticas como os fatores ambientais é crucial para enfrentar o desafio das CVDs e melhorar a saúde da população brasileira como um todo.

REFERÊNCIAS

ABRAHAM, G. et al. Genomic prediction of coronary heart disease. *European Heart Journal*, v. 37, n. 43, p. 3267–3278, 2016.

AKAI, J. et al. A novel SCN5A mutation associated with idiopathic ventricular fibrillation without typical ECG findings of Brugada syndrome. *FEBS Letters*, v. 479, n. 1–2, p. 29–34, 11 ago. 2000.

BACKBONE, G. *BeadChip*. p. 1–6, 2016.

BENZEVAL, M.; AGUIRRE, E.; KUMARI, M. Understanding Society: health, biomarker and genetic data. *Fiscal Studies*, v. 44, n. 4, p. 399–415, 1 dez. 2023.

CAVALLI-SFORZA, L. L.; FELDMAN, M. W. The application of molecular genetic approaches to the study of human evolution. *Nature Genetics*, 2003.

CHANG, C. C. ; C. C. C. ; T. L. C. ; V. S. ; P. S. M. ; L. J. J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*. v. 4, p. 7, 2015.

DE OLIVEIRA, G. M. M. et al. Cardiovascular Statistics - Brazil 2021. *Arquivos Brasileiros de Cardiologia*, v. 118, n. 1, p. 115, 2022.

ESCHER, L. M. et al. Challenges in selecting admixture models and marker sets to infer genetic ancestry in a Brazilian admixed population. *Scientific Reports*, v. 12, n. 1, p. 1–12, 2022.

FISH, A. E. et al. Local ancestry transitions modify snp-trait associations. *Pacific Symposium on Biocomputing*, v. 0, n. 212669, p. 424–435, 2018.

FUNDAÇÃO OSWALDO CRUZ. Painel de Indicadores - Pesquisa Nacional de Saúde.

HARST, P. VAN DER; VERWEIJ, N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Molecular Medicine*, p. 433–443, 2018.

HOJMAN, L.; KARSULOVIC, C. Cardiovascular Disease-Associated Skin Conditions. n. February, p. 43–53, 2022.

JIN, X. Y. et al. Biogeographic origin prediction of three continental populations through 42 ancestry informative SNPs. *Electrophoresis*, v. 41, n. 3–4, p. 235–245, 2020.

JORDAN, E. et al. Genetic Architecture of Dilated Cardiomyopathy in Individuals of African and European Ancestry. *JAMA*, v. 330, n. 5, p. 432–441, 1 ago. 2023.

KAPTOGE, S. et al. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *The Lancet Global Health*, v. 7, n. 10, p. e1332–e1345, 1 out. 2019.

KESSLER, T.; VILNE, B.; SCHUNKERT, H. The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease. *EMBO Molecular Medicine*, v. 8, n. 7, p. 688–701, jul. 2016.

KOCKUM, I.; HUANG, J.; STRIDH, P. Overview of Genotyping Technologies and Methods. *Current Protocols*, v. 3, n. 4, 1 abr. 2023.

MANSOURNIA, M. A.; NAZEMIPOUR, M.; ETMINAN, M. P-value, compatibility, and S-value. *Global Epidemiology Elsevier Inc.*, , 1 dez. 2022.

MAZZACCARA, C. et al. A common polymorphism in the SCN5A gene is associated with dilated cardiomyopathy. *Journal of Cardiovascular Medicine*, v. 19, n. 7, p. 344–350, 2018.

MENSAH, G. A. et al. Diseases in Racial and Ethnic Minority Populations. v. 125, n. 1, p. 7–13, 2020.

NIKPAY, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*, v. 47, n. 10, p. 1121–1130, 29 set. 2015.

NORRIS, E. T. et al. Genetic ancestry, admixture and health determinants in Latin America. *BMC Genomics*, v. 19, n. Suppl 8, 2018.

OLESEN, M. S. et al. Genetic Loci on Chromosomes 4q25, 7p31, and 12p12 Are Associated With Onset of Lone Atrial Fibrillation Before the Age of 40 Years. *Canadian Journal of Cardiology*, v. 28, n. 2, p. 191–195, mar. 2012.

- PARK, S. H.; HAN, K. How to Clearly and Accurately Report Odds Ratio and Hazard Ratio in Diagnostic Research Studies? *Korean Journal of Radiology* Korean Radiological Society, , 1 ago. 2022.
- PEDEN, J. F. et al. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nature Genetics*, v. 43, n. 4, p. 339–346, fev. 2011.
- PRAPIADOU, S.; DEMEL, S. L.; HYACINTH, H. I. Genetic and Genomic Epidemiology of Stroke in People of African Ancestry. v. 188, n. 2, 2018.
- RHODES, C. J. et al. Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *The Lancet Respiratory Medicine*, v. 7, n. 3, p. 227–238, 1 mar. 2019.
- SANGER INSTITUTE. HapMap.
- SCHUNKERT, H. et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*, v. 43, n. 4, p. 333–340, 2011.
- SEIFERT, M. B. et al. Genetic variants on chromosomes 7p31 and 12p12 are associated with abnormal atrial electrical activation in patients with early-onset lone atrial fibrillation. *Annals of Noninvasive Electrocardiology*, v. 24, n. 6, 1 nov. 2019.
- SHASTRY, B. S. SNPs: impact on gene function and phenotype. *Methods in molecular biology* (Clifton, N.J.), 2009.
- SINGH, S. et al. Systematic Review of Genomic Associations with Blood Pressure and Hypertension in Populations with African-Ancestry. *Frontiers in Genetics*, v. 12, n. October, 2021.
- SMITS, P. et al. The Transcription Factors L-Sox5 and Sox6 Are Essential for Cartilage Formation proliferate and abundantly produce the cartilage extra-cellular matrix. Chondroblasts then establish cartilage growth plates, in which they proceed layer by layer *Developmental Cell*. [s.l: s.n.].
- SOUZA, A. M. DE et al. A systematic scoping review of the genetic ancestry of the Brazilian population. v. 508, p. 495–508, 2019.
- TCHEANDJIEU, C. et al. *HHS Public Access*. v. 28, n. 8, p. 1679–1692, 2023.

TEMPRANO-SAGRERA, G. et al. Multi-phenotype analyses of hemostatic traits with cardiovascular events reveal novel genetic associations. *Journal of Thrombosis and Haemostasis*, v. 20, n. 6, p. 1331–1349, 1 jun. 2022.

TESLOVICH, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, v. 466, n. 7307, p. 707–713, 5 ago. 2010.

TOMA, T. T.; DAWSON, J. M.; ADJEROH, D. A. Correction to: Human ancestry identification under resource constraints -what can one chromosome tell us about human biogeographical ancestry? (*BMC Med Genomics* DOI: 10.1186/s12920-018-0412-4). *BMC Medical Genomics*, v. 12, n. 1, 2019.

VOGEL, S. et al. medicina A Higher Polygenic Risk Score Is Associated with a Higher Recurrence Rate of Atrial Fibrillation in Direct Current Cardioversion-Treated Patients. 2021.

WIMLEY, W. C. Translational Perspective on Epigenetics in Cardiovascular Disease. *Physiology & behavior*, v. 176, n. 10, p. 139–148, 2017.

YAHYA, P. et al. Ancestry-informative marker (AIM) SNP panel for the Malay population. *International Journal of Legal Medicine*, v. 134, n. 1, p. 123–134, 2020.

YU, C. et al. Detecting Genotype-Population Interaction Effects by Ancestry Principal Components. *Frontiers in Genetics*, v. 11, n. April, p. 1–12, 2020.

YU, M. et al. Genome-Wide Association Meta-Analysis Supports Genes Involved in Valve and Cardiac Development to Associate With Mitral Valve Prolapse. *Circulation: Genomic and Precision Medicine*, v. 14, n. 5, p. E003148, 1 out. 2021.