



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
MESTRADO EM SAÚDE PÚBLICA**

**IDENTIFICAÇÃO DE SNPs (*Single Nucleotide Polymorphisms*) DE
SUSCETIBILIDADE AO CÂNCER EM UMA POPULAÇÃO MISCIGENADA
LONGEVA DO NORDESTE BRASILEIRO**

STEFFANY LARISSA GALDINO GALISA

CAMPINA GRANDE – PB

2020

STEFFANY LARISSA GALDINO GALISA

**IDENTIFICAÇÃO DE SNPs (*Single Nucleotide Polymorphisms*) DE
SUSCETIBILIDADE AO CÂNCER EM UMA POPULAÇÃO MISCIGENADA
LONGEVA DO NORDESTE BRASILEIRO**

Dissertação apresentada à Universidade Estadual da Paraíba - UEPB, em cumprimento aos requisitos necessários para o título de Mestre em Saúde Pública pelo Programa de Pós-Graduação em Saúde Pública.

Orientador: Prof^o. Dr^o Mathias Weller

Co-orientadora: Prof^a. Dr^a Priscila Lima Jacob

CAMPINA GRANDE – PB

2020

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

G149i Galdino, Steffany Larissa da Silva.
Identificação de SNPs (Single Nucleotide Polymorphisms) de suscetibilidade ao câncer em uma população miscigenada longeva do nordeste brasileiro [manuscrito] / Steffany Larissa da Silva Galdino. - 2020.
75 p. : il. colorido.
Digitado.
Dissertação (Mestrado em Saúde Pública) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-Graduação e Pesquisa, 2020.
"Orientação : Prof. Dr. Mathias Weller, Departamento de Biologia - CCBS."
"Coorientação: Profa. Dra. Priscila Lima Jacob, Departamento de Biologia - CCBS."
1. Ancestralidade. 2. Câncer. 3. Single Nucleotide Polymorphisms - SNP. I. Título

21. ed. CDD 614

FOLHA DE APROVAÇÃO

Nome: Steffany Larissa Galdino Galisa

Título: Identificação de SNPs (*Single Nucleotide Polymorphisms*) de suscetibilidade ao câncer em uma população longeva miscigenada do nordeste brasileiro.

Orientador: Prof^o. Dr^o. Mathias Weller

Co-orientadora: Prof^a. Dr^a Priscila Lima Jacob

Dissertação apresentada à Universidade Estadual da Paraíba - UEPB, em cumprimento aos requisitos necessários para o título de Mestre em Saúde Pública pelo Programa de Pós-Graduação em Saúde Pública.

Orientador: Prof^o Dr^o Mathias Weller

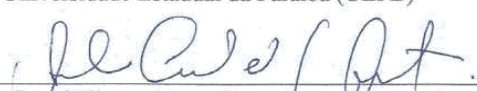
Co-orientadora: Prof^a. Dr^a Priscila Lima Jacob

Aprovada em: 27/08/2020.

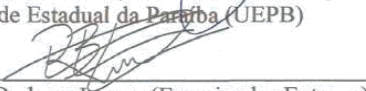
BANCA EXAMINADORA



Prof^o Dr^o Mathias Weller (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof^a. Dr^a Silvana Santos (Examinadora Interna)
Universidade Estadual da Paraíba (UEPB)



Prof^o Dr^o Renan Barbosa Lemes (Examinador Externo)
Universidade de São Paulo (USP)

AGRADECIMENTOS

A Deus que me deu saúde, força e paciência para superar as dificuldades.

A minha família e esposo pelo amor, incentivo e suporte emocional.

Ao meu orientador Mathias Weller, pelo suporte, conhecimento, incentivo, amizade e que participa mais uma vez do fim de uma etapa na minha vida acadêmica.

A minha coorientadora Priscila Lima Jacob, um obrigado especial, pelos ensinamentos, parceria e incentivo. Suas orientações me fizeram conseguir realizar este trabalho. Além de co-orientadora você se tornou uma grande amiga. Ao grupo de pesquisa Epidemiologia e Genética de Câncer pelo conhecimento proporcionado.

As minhas amigas de mestrado Adriana, Radmila e Isa, obrigada pela amizade, esses dois anos foram muito especiais ao lado de vocês.

A Fundação de Apoio à Pesquisa do Estado da Paraíba – FAPESQ, pelo apoio financeiro.

“A menos que modifiquemos à nossa maneira de pensar, não seremos capazes de resolver os problemas causados pela forma como nos acostumamos a ver o mundo”.

(Albert Einstein)

RESUMO

O câncer é a principal causa de morbidade e mortalidade no mundo, sendo considerado um fator limitante para a expectativa de vida. Com o envelhecimento, observa-se maior incidência de câncer ocasionada por alterações fisiológicas e genéticas relacionadas com a idade. Os polimorfismos de nucleotídeo único (SNPs), alterações genéticas mais comuns no genoma humano, podem contribuir para a suscetibilidade ao câncer, atuando como biomarcadores. O objetivo do trabalho foi investigar na população miscigenada de longevos de Brejo dos Santos, variantes genéticas de suscetibilidade ao câncer e realizar um estudo de ancestralidade. Foram selecionados 73 idosos saudáveis com ≥ 80 anos, residentes no município de Brejo dos Santos - PB. Realizou-se a genotipagem dos idosos utilizando a plataforma de SNP array *Axiom® Genome-Wide LAT 1 Array Plate*. Foi utilizado o *Genotyping Console™ software (Version 4.2; Affymetrix Inc.)* para análise dos genótipos e controle de qualidade, aplicando os filtros: *SNP call rate >97%*; *Hardy-Weinberg test (HWE)* com *p-valor > 0,05*, e os genes *BRCA1*, *BRCA2*, *TP53*, *AURKA*, *CCND1*, *CDKN1A*, *ATM*, *ERCC5*, *XRCC1*, *MC1R*, *VEGF*, *MMP7*, *ERCC1*, *ERCC2*, *RBI*, *HNF1B*, *NCOA3*, *MLH1*, *MITF* e *CDH1*. Foram identificadas 90 variantes genéticas de suscetibilidade ao câncer, sendo que 47,25% das variantes apresentaram ancestralidade europeia, 49,45% ancestralidade europeu/africano e 3,29% das variantes ancestralidade europeu/nativo-americano. Foram observadas diferenças na frequência alélica entre populações de regiões geográficas diferentes como Paraíba e São Paulo e em populações mundiais. Dos 48 haplótipos identificados na população, 23 apresentaram apenas alelos de risco para o câncer, cinco apresentaram alelos que aumentam e diminuem o risco e cinco apresentaram apenas alelos protetores para o câncer. As maiores frequências observadas na população foram de haplótipos europeus e haplótipos europeu/africano com 45,83% cada. Haplótipos de origem africana tiveram uma frequência de 4,16%, e haplótipos europeu/nativo-americano e europeu/africano/nativo-americano tiveram uma frequência de 2,08% cada. Neste trabalho, buscou-se combinar a análise de haplótipos em grupos miscigenados para explorar a arquitetura genética do câncer na população. Os resultados fornecem o primeiro relatório sobre polimorfismos genéticos comuns de câncer em uma população consanguínea do nordeste brasileiro, que podem estabelecer um banco de dados de base genética para orientar futuros estudos de associação.

Palavras-chave: Câncer, SNPs, ancestralidade.

ABSTRACT

Cancer is the main cause of morbidity and mortality in the world, being considered a limiting factor for life expectancy. With aging, there is a higher incidence of cancer caused by physiological and genetic changes related to age. Single nucleotide polymorphisms (SNPs), the most common genetic changes in the human genome, can contribute to cancer susceptibility, acting as biomarkers. The objective of the work was to investigate the genetic mix of susceptibility to cancer in the long-lived population of Brejo dos Santos and carry out a study of ancestry. 73 healthy elderly people aged ≥ 80 years, living in the municipality of Brejo dos Santos - PB, were selected. The elderly were genotyped using the Axiom® Genome-Wide LAT 1 Array Plate SNP array platform. The Genotyping Console™ software (Version 4.2; Affymetrix Inc.) was used for genotype analysis and quality control, applying the filters: SNP call rate $> 97\%$; Hardy-Weinberg test (HWE) with p -value > 0.05 , and the BRCA1, BRCA2, TP53, AURKA, CCND1, CDKN1A, ATM, ERCC5, XRCC1, MC1R, VEGF, MMP7, ERCC1, ERCC2, RB1, HNF1B genes, NCOA3, MLH1, MITF and CDH1. Ninety genetic variants of cancer susceptibility were identified, with 47.25% of the variants having European ancestry, 49.45% European / African ancestry and 3.29% of the European / Native American ancestry variants. Differences in allele frequency were observed between populations from different geographic regions such as Paraíba and São Paulo and in world populations. Of the 48 haplotypes identified in the population, 23 had only cancer risk alleles, five had risk increasing and decreased alleles, and five had only cancer protective alleles. The highest frequencies observed in the population were European haplotypes and European / African haplotypes with 45.83% each. Haplotypes of African origin had a frequency of 4.16%, and haplotypes European / Native American and European / African / Native American had a frequency of 2.08% each. In this work, we sought to combine the analysis of haplotypes in miscegenated groups to explore the genetic architecture of cancer in the population. The results provide the first report on common genetic polymorphisms of cancer in an inbred population of northeastern Brazil, which can establish a genetic-based database to guide future association studies.

Keywords: Cancer, SNPs, ancestry.

LISTA DE FIGURAS

Figura 1- Frequência alélica menor (MAF) das variantes genéticas nas populações da Paraíba e São Paulo.	26
Figura 2 - Porcentagem de SNPs que apresentaram o alelo menor mais frequente nas populações do 1000 Genomas e população brasileira (Brejo dos Santos).	27
Figura 3- Frequência e origem do alelo maior e alelo menor dos 90 SNPs identificados na população de idosos de Brejo dos Santos.	28
Figura 4 - Gráfico de Desequilíbrio de ligação e frequências de haplótipos.....	30
Figura 5 - Origem ancestral dos haplótipos por cromossomo.....	34
Figura 6 - Distribuição dos haplótipos e suas origens identificados nos idosos longevos de Brejo dos Santos.	35

LISTA DE TABELAS

Tabela 1- Frequência alélica, localização, função genética e significado clínico das 90 variantes genéticas identificadas nos 73 longevos de Brejo dos Santos.	22
--	----

LISTA DE SIGLAS

ABraOM	Arquivo Brasileiro Online de Mutações
<i>ATM</i>	Ataxia telangiectasia mutated
<i>AURKA</i>	Aurora kinase A
<i>BRCA1</i>	Breast cancer 1
<i>BRCA2</i>	Breast cancer 2
<i>CCND1</i>	Cyclin D1
<i>CDH1</i>	E-cadherin
<i>CDKN1A</i>	Cyclin Dependent Kinase Inhibitor 1A
CEGH-CEL	Centro de Pesquisa sobre o Genoma Humano e Células-Tronco
<i>DNA</i>	Deoxyribonucleic acid
<i>ERCC1</i>	Excision Repair 1
<i>ERCC2</i>	Excision Repair 2
<i>ERCC5</i>	Excision Repair 5
GWAS	Genome-wide association study
<i>HNF1B</i>	Hepatocyte nuclear factor 1-beta
<i>HWE</i>	Hardy-Weinberg equilibrium
LAI	Local-ancestry inference
LD	Linkage disequilibrium
<i>MAF</i>	Minor Allele Frequency
<i>MC1R</i>	Melanocortin 1 receptor
<i>MITF</i>	Melanocyte inducing transcription factor
<i>MLH1</i>	<i>Human mutL homolog 1</i>
<i>MMP7</i>	Matrix metalloproteinase 7
<i>NCOA3</i>	<i>Nuclear receptor coactivator 3</i>
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana de Saúde
<i>RB1</i>	RB Transcriptional Corepressor 1
SABE	Estudo Saúde, Bem-Estar e Envelhecimento
<i>SNP</i>	Single nucleotide polymorphism
<i>TP53</i>	Tumor protein p53
UEPB	Universidade Estadual da Paraíba
USP	Universidade de São Paulo

VEGF

Vascular endothelial growth factor

XRCC1

X-ray repair cross complementing 1

Sumário

1 INTRODUÇÃO	13
2 OBJETIVOS	17
2.1 Objetivo Geral	17
2.2 Objetivos Específicos	17
3 MÉTODOS	18
3.1 População de estudo	18
3.2 Extração de DNA	18
3.3 Genotipagem e controle de qualidade	18
3.4 Frequência alélica e genotípica.....	19
3.5 Inferência de Ancestralidade Local (LAI).....	20
3.6 Análise de dados de genotipagem e análise de ligação	21
3.7 Origem dos haplótipos.....	21
4 RESULTADOS	21
4.1 Frequência alélica das variantes genéticas	21
4.2 Frequência genotípica das variantes genéticas	25
4.3 Comparação das variantes de Brejo dos Santos e São Paulo	25
4.4 Frequência alélica menor (MAF) das variantes genéticas nas populações mundiais	27
4.5 Inferência de Ancestralidade Local	27
4.6 Haplótipos.....	28
4.7 Origem dos haplótipos.....	33
4.8 Haplótipos e câncer	35
5 DISCUSSÃO	36
6 CONCLUSÃO	43
REFERÊNCIAS	44

APÊNDICES	57
APÊNDICE A - Revisão de literatura dos 90 SNPs identificados nos 73 idosos longevos de Brejo dos Santos.	58
APÊNDICE B - Frequência genotípica dos 90 SNPs e o p-valor de Hardy Weinberg. 65	
APÊNDICE C - Frequência do alelo menor (MAF) para os 90 SNPs nas populações do 1000 Genomas (Europeu, Africano, Nativo-americano miscigenado) e população de Brejo dos Santos.	69
APÊNDICE D - Média e desvio-padrão das MAFs nas populações mundiais (Europeu, Africana e Nativo-americano miscigenado) e população de Brejo dos Santos.	72
APÊNDICE E – Haplótipos, ancestralidade e referência dos alelos de haplótipos que modificam o risco de câncer.	73

1 INTRODUÇÃO

A expectativa de vida humana ao longo do século XX era de aproximadamente 50 a 60 anos e, nas décadas seguintes, foi crescendo consideravelmente em todo o mundo, atingindo cerca de 80 anos de idade em muitos países desenvolvidos (SMETANA et al., 2016). Com o envelhecimento populacional e o aumento da expectativa de vida, o subgrupo dos idosos longevos (acima de 80 anos de idade) tem sido o segmento populacional que mais cresce mundialmente. Nos países desenvolvidos, usa-se a expressão “oldest-old”, considerando como critério a idade acima de 85 anos, já que a expectativa de vida é maior nesses países (FA et al., 2015). No Brasil, a população de idosos acima de 65 anos era de 14 milhões em 2012; tendo sido duplicada em 2017 para 28 milhões e, em 2027, estima-se que sejam 38,5 milhões (IBGE, 2017). Dentre os idosos, há uma tendência ao aumento da prevalência de idosos longevos; os quais representavam, em 2012, 1,6% do grupo (LOURENCO et al., 2012).

O aumento da expectativa de vida relacionada aos avanços nos cuidados de saúde, controle de higiene e disponibilidade de alimentos trouxe, conseqüentemente, uma proporção crescente de pessoas em idade suscetível ao câncer (AUNAN; CHO; SØREIDE, 2017). O crescente número de novos casos de câncer em todo o mundo, acarreta implicações importantes na área da saúde, sendo uma barreira importante para aumentar a expectativa de vida em todos os países no século XXI. Em 2015, o câncer foi a primeira ou segunda causa de morte antes dos 70 anos em 91 dos 172 países, e ocupa o terceiro ou quarto lugar em outros 22 países (WHO, 2018).

O envelhecimento humano e o desenvolvimento do câncer compartilham vários mecanismos em comum, tais como, instabilidade genômica, atrito de telômeros, alterações epigenéticas, diminuição da detecção de nutrientes, metabolismo alterado e senescência celular. Além disso, ocorre a redução da atividade de máquinas de reparo do DNA fundamental para a manutenção da integridade do genoma. As células envelhecidas são hipoativas com mutações desvantajosas acumuladas, incapacidade de divisão celular e apresentam uma capacidade reduzida de produção e consumo de energia. Dessa forma, o envelhecimento e o câncer estão intimamente interconectados, sendo por isso considerado uma doença do envelhecimento (AUNAN; CHO; SØREIDE, 2017; SMETANA et al., 2016).

O câncer é um grupo de doenças caracterizadas pelo crescimento celular não regulado e pela invasão e disseminação de células do local de origem, ou local primário,

para outros locais do corpo (PECORINO, 2016). É considerado uma das doenças mais importantes do mundo e é complicado por ser multifatorial em uma visão epidemiológica (GHONCHEH, POURNAMDAR, SALEHINIYA, 2016). O processo canceroso é resultado da função celular perturbada. Isso ocorre devido ao acúmulo de muitas alterações genéticas e epigenéticas dentro da célula, expressas pelo acúmulo de aberrações cromossômicas ou moleculares, o que leva à instabilidade genética (LEWANDOWSKA et al., 2018).

A maior incidência de câncer em idosos é atribuída ao acúmulo de mutações somáticas adquiridas e alterações epigenéticas durante a vida, que podem ser aceleradas por fatores ambientais. Projetos de sequenciamento de genoma de câncer em larga escala confirmaram que a maioria dos cânceres abriga algumas mutações somáticas altamente recorrentes em oncogenes clássicos e genes supressores de tumores juntamente com mutações somáticas raras. Há também centenas de variantes germinativas de alto impacto, comuns e raras, que são responsáveis pela herdabilidade do câncer (QING et al., 2020). Uma pequena proporção (5-10%) dos cânceres é causada por mutações hereditárias nos genes (ASHTON-PROLLA, SEUANEZ, 2016).

Existem várias classes de genes que estão envolvidas nesse processo: genes supressores de tumor (*BRCA1*, *BRCA2*, *TP53*, *CDH1*, *ATM*, *RBI*, *MLH1*), oncogenes (*AURKA*), pronto-oncogene (*CCND1*), genes envolvidos no reparo do DNA (*XRCC1*, *ERCC1*, *ERCC2* e *ERCC5*, *MLH1*) e controle do ciclo celular (*CDKN1A*) (DUFLOTH et al., 2005; FETT-CONTE; SALLES, 2002; GUO et al., 2018; ALAO, 2007; BERRY et al., 2019; DUTTO et al., 2015, KAGUE et al., 2010; BOSE et al., 2009; GÓMEZ-DÍAZ et al., 2015; SOLTYS et al., 2013; MARTINEZ et al., 2009; PARK et al., 2002).

Uma das estratégias utilizadas para estudar a suscetibilidade ao câncer é o uso de polimorfismos de nucleotídeo único (SNPs). O polimorfismo de nucleotídeo único refere-se a uma variação de base única na sequência genômica, consistindo em substituições, de nucleotídeo único, além de ser considerada a variação mais comum no genoma humano. Os SNPs podem resultar em diferentes produtos funcionais, afetando a suscetibilidade individual a doenças, como o câncer. Assim, podem ser considerados como biomarcadores para prever o risco de tumores; entretanto algumas mutações podem não expressar tumores devido efeitos de estratificação populacional (WEN; XU; YUAN, 2018; ZHOU et al., 2018). Compreender os mecanismos subjacentes aos efeitos dos SNPs que resultam em suscetibilidade ao câncer é fundamental para entender a patogênese molecular de vários tipos de câncer (DENG et al., 2017).

Os SNPs podem ser reunidos em haplótipos, que são blocos quebrados de forma não-aleatória por recombinação ou alelos que estão em desequilíbrio de ligação (HAMDI et al., 2018). A caracterização de haplótipos em populações humanas possui um impacto na epidemiologia molecular, assim como implicações importantes na compreensão da biologia do câncer (ERICHSEN; CHANOCK, 2004). A frequência desses haplótipos pode variar entre as populações humanas e essas diferenças podem ser utilizadas para o mapeamento de variantes de risco para determinadas doenças (HAMDI et al., 2018), além de ser uma forma importante para capturar a diversidade genética em populações humanas (ERICHSEN; CHANOCK, 2004). Outras aplicações referem-se à definição do risco genético e das trajetórias demográficas de mutações genéticas associadas ao câncer em populações atuais e ancestrais.

A história demográfica humana se reflete em padrões específicos de mutações compartilhadas entre os genomas de diferentes populações. Os genomas abrigam informações ricas sobre a história da população, codificadas em padrões de mutações e recombinações (WANG et al., 2020). Dessa forma, a história demográfica pode afetar a diversidade genética das populações humanas de maneira diferente, sendo a inferência da ancestralidade local uma ferramenta refinada e indispensável para o estudo da origem das mutações genéticas, associadas a doenças como o câncer em diferentes cenários demográficos. As informações de ancestralidade local podem ajudar na identificação de variantes causais, na origem de doenças genéticas por haplótipos e frequência alélica, e na datação do surgimento de doenças genéticas raras e complexas (DE FARIAS et al., 2018).

A miscegenação ocorre quando duas ou mais populações previamente independentes se cruzam, resultando na introdução de novas linhagens genéticas, como ocorreu em latino-americanos que é uma população miscegenada (SJAARDA et al., 2020). A população brasileira é caracterizada por ser uma população tri-híbrida, isto devido às contribuições genéticas de nativos americanos, europeus e africanos, além disso, a contribuição das três diferentes populações ancestrais é diferente entre as regiões brasileiras, ou seja, cada região vai ter sua particularidade étnica (FARIAS et al., 2018; SECOLIN et al., 2019). As populações miscigenadas permitem identificar variantes genéticas relacionadas a proteção e suscetibilidade que podem ter sido conservadas durante o isolamento de populações nativo-americanas por quase dez mil anos, e o posterior influxo de escravos africanos e invasores europeus nos últimos 520 anos.

No Nordeste do Brasil, que abriga 28% da população brasileira, existem muitas comunidades pequenas, localizadas no sertão que historicamente eram isoladas ou semi-isoladas geograficamente, apresentando altos níveis de consaguinidade, com frequência total cerca de 15 vezes maior que na região do sul (FARIAS et al., 2018; WELLER et al., 2012). No sertão da Paraíba, a população do município de Brejo dos Santos apresenta uma taxa relativamente alta de consanguinidade, no qual 18% dos casamentos são entre indivíduos aparentados (WELLER et al., 2012; LIMA et al., 2019), contribuindo na incidência de doenças monogênicas, como as que predisõem ao câncer.

A população de idosos longevos de Brejo dos Santos faz parte do Estudo Saúde, Bem-Estar e Envelhecimento (SABE) - Paraíba. O SABE foi inicialmente idealizado pela Organização Pan-Americana de Saúde/Organização Mundial da Saúde (OPAS/OMS) como um inquérito multicêntrico para traçar o perfil das condições de vida e saúde das pessoas idosas em sete centros urbanos na América Latina e Caribe. No Brasil, o projeto teve início em 2000, no qual foi desenvolvido na cidade de São Paulo, por apresentar a maior população idosa e diversificada no país (LEBRÃO et al., 2019). Em 2014, a proposta do SABE foi ampliada para municípios do nordeste a partir de uma parceria entre a UEPB e USP.

A análise de variantes genéticas e haplótipos se torna particularmente interessante de se estudar na população de Brejo dos Santos, principalmente pelo fato de ser uma população longeva, que não desenvolveram câncer ao longo da vida e consanguínea possuindo alto grau de homogeneidade genética. Como o risco a doença varia significativamente entre diferentes populações, a identificação de variantes genéticas e diferenças genético-populacionais no genoma humano, torna possível analisar a predisposição genética e o impacto da ancestralidade a certas doenças, como o câncer.

Devido à lacuna existente no estudo de ancestralidade local de variantes genéticas associadas ao câncer em populações miscigenadas brasileiras, foi realizada a análise de variantes genéticas de suscetibilidade do câncer na população de Brejo dos Santos. Os haplótipos e as frequências alélicas foram estimadas a partir de populações ancestrais e foi realizada a identificação do risco da presença de diversas variantes genéticas associadas ao câncer na população miscigenada de longevos.

2 OBJETIVOS

2.1 Objetivo Geral

- Identificar SNPs de suscetibilidade ao câncer na população miscigenada de idosos longevos no município de Brejo dos Santos na Paraíba.

2.2 Objetivos Específicos

- Rastrear SNPs que já foram previamente associados a suscetibilidade de diversos tipos de cânceres presentes nos genes *BRCA1*, *BRCA2*, *TP53*, *AURKA*, *CCND1*, *CDKN1A*, *ATM*, *XRCC1*, *ERCC5*, *MC1R*, *VEGF*, *MMP7*, *ERCC1*, *ERCC2*, *RBI*, *HNF1B*, *NCOA3*, *MLH1*, *MITF* e *CDH1*.
- Relatar as frequências alélicas das variantes genéticas (SNPs) encontradas.
- Relatar as frequências genotípicas das variantes e verificar se estão em equilíbrio de Hardy-weinberg.
- Comparar a frequência do alelo menor nos idosos longevos do SABE-PB e SABE-SP.
- Comparar a frequência do alelo menor nos idosos de Brejo dos Santos na Paraíba com a populações do Banco de dados 1000 Genomas.
- Inferir a ancestralidade local dos indivíduos para as variantes genéticas estudadas.
- Investigar os padrões de desequilíbrio de ligação nos genes citados na população.
- Estimar a frequência de haplótipos relacionada às variantes associadas aos cânceres.
- Identificar a origem dos haplótipos a partir do estudo de ancestralidade local, avaliando o efeito da ancestralidade na população.

3 MÉTODOS

3.1 População de estudo

Do total de 188 idosos com 80 anos ou mais de Brejo dos Santos (PB), foram amostrados 73 idosos saudáveis para genotipagem (35 homens e 38 mulheres). Todos os participantes assinaram o Termo de Consentimento Livre e Esclarecido e o projeto foi aprovado pelo Comitê de Ética em Pesquisa Envolvendo Seres Humanos da Universidade Estadual da Paraíba (UEPB), com o número do protocolo n° 0359.0.133.000-11.

3.2 Extração de DNA

A extração do DNA genômico foi realizada na UEPB e no Centro de Pesquisa sobre o Genoma Humano e Células-Tronco, CEGH-CEL da Universidade de São Paulo – USP. Foram coletados 5 ml de sangue periférico de cada indivíduo para extração de DNA. O DNA foi extraído utilizando o kit de extração, as amostras foram quantificadas com o uso do espectrofotômetro Nanodrop ND-1000 (*Thermo Scientific, Massachusetts, EUA*), as alíquotas foram diluídas em água *milli-Q* e armazenadas em condições apropriadas no CEGH-CEL da USP.

3.3 Genotipagem e controle de qualidade

A genotipagem dos 73 idosos com 80 anos ou mais foi realizada utilizando a plataforma *Axiom® Genome-Wide LAT 1 Array (Affymetrix, USA)*, no CEGH-CEL da USP seguindo as recomendações do fabricante (*Affymetrix, USA*). Baseado nos arquivos (.cel e .arr), obtidos pelo microarranjo, foi realizada análise da genotipagem pelo software *Genotyping Console™ (Version 4.2, Affymetrix Inc.)*, que analisa os polimorfismos de nucleotídeo único (SNPs), detecção de *indels* e variação de número de cópias (CNV) dos indivíduos.

Os genes desencadeadores do câncer foram identificados em estudos publicados no PubMed de 2005 a 2020. Os termos de pesquisa: "Oncogenes AND tumor suppressor gene" e "genes AND cancer". Os estudos elegíveis foram aqueles que relataram principalmente a alelos de maior ou menor suscetibilidade ao câncer, associados também prognóstico e sobrevida ao câncer e a haplótipos. Os 20 genes selecionados foram selecionados: *BRCA1, BRCA2, TP53, AURKA, CCND1, CDKN1A, ATM, ERCC5,*

XRCC1, MC1R, VEGF, MMP7, ERCC1, ERCC2, RB1, HNF1B, NCOA3, MLH1, MTF e CDH1.

Como controle de qualidade inicial, foram utilizados os seguintes filtros: $DQC \geq 0,82$ (*Dish quality control*) $QC\ call\ rate\ (Quality\ control\ call\ rate) \geq 92$, $Average\ call\ rate\ for\ passing \geq 97$ e $Minor\ allele\ cutoff \geq 2$. Para análise de SNPs previamente associados ao câncer na literatura, novos filtros foram utilizados: genes (*BRCA1, BRCA2, TP53, AURKA, CCND1, CDKN1A, ATM, ERCC5, XRCC1, MC1R, VEGF, MMP7, ERCC1, ERCC2, RB1, HNF1B, NCOA3, MLH1, MTF e CDH1*), $SNP\ call\ rate >97\%$ e *Hardy-Weinberg test* (HWE) com p-valor maior do que 0,05 e igual a 1,00.

Do total de 805.712 SNPs, o número de polimorfismos que abrigam os genes do câncer na matriz LAT foi de 2.948, dos quais 90 SNPs foram selecionados para a análise e classificados com base no navegador do genoma Ensembl como: variante intrônica, variantes a *upstream* e *downstream*, regiões não traduzidas (3'-UTR; 5'-UTR), variantes sinônimas ou não-sinônimas. Não foram identificadas variantes de suscetibilidade ao câncer na matriz LAT nos genes *MC1R, MLH1* e *MTF* restando 17 genes para análise de frequência de alelos.

3.4 Frequência alélica e genotípica

Foram identificadas a frequência alélica maior (alelo mais comum) e menor (alelo menos comum), posição física (hg19), função genética, classificação e significado clínico dos SNPs. A significância clínica das variantes foi obtida através do banco de dados Clinvar (<https://www.ncbi.nlm.nih.gov/clinvar/>) que fornece relações entre variantes genéticas humanas e as interpretações de seu significado para a doença Clinvar.

A frequência genotípica e o *p-value* do equilíbrio de Hardy-Weinberg das variantes do na população estudada foram obtidas pelo arquivo *SNP Summary Table* de cada gene pelo programa *Genotyping ConsoleTM*. As variantes genéticas foram verificadas se estavam em equilíbrio de Hardy-Weinberg. Testar o equilíbrio de Hardy-Weinberg (HWE) é um componente importante em quase todas as análises de dados genéticos populacionais, servindo também como uma verificação da qualidade dos dados (HAO; STOREY et al., 2019).

O alelo menor é o alelo menos frequente na população. Estimar a frequência alélica menor (MAF), é particularmente importante na genética de populações, pois representam a maior parte da variabilidade nas populações naturais, diferenciando variantes comuns e raras (RAINERI et al., 2012; KIM et al., 2011). Além disso, os alelos

menores exibem uma tendência maior a serem alelos de risco, principalmente quando a frequência do alelo menor é inferior a 0,1 (KIDO et al., 2018).

A frequência alélica menor (MAF) das variantes identificadas na população estudada na Paraíba (SABE-PB) foi comparada com o banco de dados sobre variantes genéticas de população idosa de São Paulo (SABE-SP) com objetivo de observar como ocorre a distribuição da frequência alélica em duas populações de idosos longevos que residem em regiões geográficas diferentes. As informações da MAF para a população de São Paulo foram obtidas no Banco de dados do Arquivo Brasileiro Online de Mutações (ABraOM) (NASLAVSKY et al., 2017).

Adicionalmente, a frequência alélica menor (MAF) das variantes estudadas foi identificada nas populações de referência européias (N = 1006), africanas (N = 1322) e nativo americano miscigenada (N = 694) no catálogo de variação genética humana - 1000 Genomes Project (1000 GENOMES PROJECT CONSORTIUM et al., 2012) presente no dbSNP (banco de dados de polimorfismos de nucleotídeo único) usando GRCh37 / hg19 como referência do genoma humano com o objetivo de identificar diferenças na distribuição alélica entre as populações mundiais.

3.5 Inferência de Ancestralidade Local (LAI)

A inferência de ancestralidade local foi executada pelo programa RFMix para determinar a origem do alelo maior e alelo menor das variantes genéticas nos genes (*BRCA1*, *BRCA2*, *TP53*, *AURKA*, *CCND1*, *CDKN1A*, *ATM*, *ERCC5*, *XRCC1*, *MC1R*, *VEGF*, *MMP7*, *ERCC1*, *ERCC2*, *RBI*, *HNFB1B*, *NCOA3* e *CDH1*). A inferência de ancestralidade local (LAI) foi estimada usando o software RFMix (MAPLES et al., 2013). O RFMix é uma abordagem discriminativa que usa campos aleatórios condicionais com base nos dados do painel de referência [10.1016/j.ajhg.2013.06.020](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44154). Foi utilizado 0,2 cM, duas iterações de maximização de expectativa, com a opção PopPhased baseada no algoritmo padrão de avanço / retrocesso. Além disso, foi utilizado o banco de dados 1KGP, incluindo iorubá, representando africanos, ibéricos representando europeus e peruanos representando nativos americanos miscigenado, conforme descrito anteriormente (FARIAS et al., 2018).

3.6 Análise de dados de genotipagem e análise de ligação

Após a análise da genotipagem pelo software *Genotyping Console*TM (Version 4.2, *Affymetrix Inc.*), os dados genéticos de todos os idosos do SABE-PB (.cel e .arr) foram exportados para o PLINK 1.9.

A partir dos arquivos binários (*.BDS), foram utilizados filtros para as variantes específicas de cada cromossomo, obtendo arquivos com extensão (*.BED e *.BIM). Posteriormente, os arquivos *.MAP e *.PED foram transformados em arquivos *.INFO e *.PED para cada cromossomo.

Os arquivos de dados dos marcadores e genótipos (INFO e PED) foram carregados no software *Haploview 4.0*. A partir disso, foram identificados blocos de haplótipos, calculando o LD pareado para cada par de variante da sequência, para os 73 idosos genotipados. A estrutura de desequilíbrio de ligação pareada (LD) foi construída com todos os SNPs avaliados para cada cromossomo. Além disso, os haplótipos foram estimados usando SNPs com $MAF \geq 5\%$.

Foram analisados os blocos de haplótipos existentes na população e os padrões de desequilíbrio de ligação, determinadas as frequências dos haplótipos para verificar o grau de compartilhamento na população e foram calculados os coeficientes de correlação. Os haplótipos com frequência >20% foram considerados como “mais frequentes”; haplótipos com frequência <5% foram considerados “haplótipos raros” e haplótipos com frequência entre 5-20% considerados “haplótipos comuns”. Além disso, os tamanhos dos haplótipos foram analisados no intuito verificar o grau de compartilhamento desses haplótipos na população.

3.7 Origem dos Haplótipos

Foi realizada a inferência de ancestralidade para os haplótipos identificados na população estudada. A inferência de ancestralidade local (LAI) foi estimada usando o software RFMix (MAPLES et al., 2013).

4 RESULTADOS

4.1 Frequência alélica das variantes genéticas

No presente estudo, foi investigado a arquitetura genética do câncer em um conjunto de 73 indivíduos da população de Brejo dos Santos. Das 2.948 variantes

genéticas identificadas, 90 já foram previamente associadas a suscetibilidade de câncer, segundo a literatura (APÊNDICE A).

A tabela 1 resume informações de posição cromossômica, frequências alélicas, função gênica e classificação de patogenicidade (Clinvar) das variantes genéticas nos genes: *BRCA1*, *BRCA2*, *TP53*, *AURKA*, *CCND1*, *CDKN1A*, *ATM*, *ERCC5*, *XRCC1*, *VEGF*, *MMP7*, *ERCC1*, *ERCC2*, *RBI*, *HNF1B*, *NCOA3* e *CDH1*. Os genes *MLH1*, *MITF* e *MC1R* não apresentaram variantes genéticas associadas à suscetibilidade de câncer.

Dos 90 SNPs selecionados (GRCh37 - hg19), 38 foram localizados em regiões intrônicas entre exons e dois em regiões intergênicas; três eram variantes de exons de transcrição não codificantes; oito eram substituições sinônimas e 21 não-sinônimas (missense); duas variantes (stop gain), três variantes *downstream*, seis *upstream*, três em 5'-UTR e quatro em 3'-UTR (Tabela 1). Dos 90 SNPs encontrados na população de Brejo dos Santos, 59 variantes tinham alelos idênticos aos descritos na literatura, sete tinham pelo menos um alelo em comum e 23 tinham alelos diferentes. A literatura não relata os alelos para uma variante (APÊNDICE A).

As variantes do *BRCA1* (rs16942, rs1799950, rs4986852, rs1799949, rs4986764, rs4986765, rs2048718 e rs799923) e do *BRCA2* (rs4987117, rs144848, rs4987047 e rs206119) foram classificadas como benignas ou provavelmente benignas para o câncer de mama e ovário de acordo com Clinvar (Tabela 1). As variantes rs1042522 (*TP53*), rs1801270 (*CDKN1A*), rs4987943, rs3092856 (*ATM*) foram classificadas como benignas ou provavelmente benignas para o câncer em geral. A variante rs2909430 do *TP53* foi classificada como benigna, mas não cita nenhuma patologia. Quatro variantes do gene *ERCC5* e uma variante do *ERCC2* foram classificadas como benignas ou provavelmente benignas para xeroderma pigmentoso. Até o momento, as demais variantes genéticas selecionadas não foram relatadas e sua importância clínica não foi fornecida.

Tabela 1- Frequência alélica, localização, função genética e significado clínico das 90 variantes genéticas identificadas nos 73 longevos de Brejo dos Santos.

Gene Crom.	SNP	Posição Física	Alelo Maior	Freq.	Alelo Menor	Freq.	Função	Significado Clínico
<i>BRCA1</i> (Crom. 17)	rs16942	41244000	T	0,7054	C	0,2945	NS (Mi)	Benigno
	rs1799950	41246481	T	0,9178	C	0,0821	NS (Mi)	Benigno
	rs4986852	41244429	C	0,9931	T	0,0068	NS (Mi)	Benigno
	rs1799949	41245466	G	0,7054	A	0,2945	S	Benigno

	rs4986764	59763347	G	0,5684	A	0,4315	NS (Mi)	Benigno
	rs4986765	59763465	T	0,6506	C	0,3493	S	PB
	rs2048718	59940819	C	0,5675	T	0,4315	5'UTR	PB
	rs799923	41251931	G	0,8151	A	0,1849	Intron	Benigno
<i>BRCA2</i> (Crom. 13)	rs4987117	32914236	C	0,9863	T	0,0136	NS (Mi)	Benigno
	rs144848	32906729	A	0,6917	C	0,3082	NS (Mi)	Benigno
	rs4987047	32953529	A	0,9718	T	0,0281	NS (Mi)	Benigno
	rs206119	32889968	G	0,7124	A	0,2876	Intron	Benigno
<i>TP53</i> (Crom. 17)	rs1042522	7579472	G	0,6901	C	0,3098	NS (Mi)	Benigno
	rs2909430	7578645	C	0,8288	T	0,1712	Intron	Benigno
	rs2078486	7583083	G	0,8699	A	0,1301	Intron	N/R
	rs2287497	7592780	G	0,7362	A	0,2638	Intron	N/R
	rs12951053	7577407	A	0,8836	C	0,1164	Intron	N/R
<i>AURKA</i> (Crom.20)	rs1047972	54961463	T	0,8356	C	0,1643	NS (Mi)	N/R
	rs6024836	54944068	G	0,6389	A	0,3611	Down	N/R
<i>CCND1</i> (Crom.11)	rs3862792	69462856	C	0,9305	T	0,0694	S	N/R
	rs1944129	69453935	C	0,5206	T	0,4794	Up	N/R
	rs614367	69328764	C	0,8836	T	0,1164	Up	N/R
<i>CDKN1A</i> (Crom.6)	rs1801270	36651971	C	0,8959	A	0,1041	NS (Mi)	Benigno
	rs2395655	36645696	A	0,5417	G	0,4583	NS (Mi)	N/R
	rs1321311	36622900	C	0,7946	A	0,2054	Intergen	N/R
<i>ATM</i> (Crom.11)	rs4987943	108121446	A	0,9862	G	0,0138	S	PB
	rs3092856	108159732	C	0,9864	T	0,0136	NS (Mi)	PB
<i>XRCC1</i> (Crom.19)	rs1799782	44057574	G	0,9521	A	0,0479	NS (Mi)	N/R
	rs25489	44056412	C	0,9453	T	0,0547	NS (Mi)	N/R
	rs25487	44055726	T	0,6987	C	0,3013	NS (Mi)	N/R
	rs2307177	44047825	T	0,9795	G	0,0205	NS (Mi)	N/R
	rs25486	44055898	C	0,6806	T	0,3194	Intron	N/R
	rs762507	44058098	T	0,6096	C	0,3904	Intron	N/R
<i>ERCC5</i> (Crom. 13)	rs17655	103528002	G	0,6507	C	0,3493	NS (Mi)	Benigno
	rs2227869	103515085	G	0,9247	C	0,0753	NS (Mi)	PB
	rs4150386	103527528	A	0,9237	C	0,0763	Intron	N/R
	rs2094258	103496759	C	0,8288	T	0,1712	Intron	N/R
	rs2296148	103498545	C	0,9042	T	0,0958	5'UTR	PB
	rs4150351	103522967	A	0,9521	C	0,0479	Intron	N/R
	rs4150360	103524762	C	0,6929	T	0,3071	Intron	Benigno
	rs4150383	103527230	G	0,7917	A	0,2083	Intron	N/R

	rs4150393	103528658	A	0,9658	G	0,0342	Up	N/R
<i>VEGF</i> (Crom.6)	rs1005230	43736496	C	0,6439	T	0,3561	Up	N/R
	rs25648	43738977	C	0,9042	T	0,0958	S	N/R
	rs3025035	43751359	C	0,8905	T	0,1095	NCtrans	N/R
	rs3025039	43752536	C	0,7917	T	0,2083	3'UTR	N/R
	rs10434	43753212	A	0,6781	G	0,3219	3'UTR	N/R
	rs3025040	43753051	C	0,7877	T	0,2123	3'UTR	N/R
	rs833052	43723335	C	0,9316	A	0,0684	Intergen	N/R
<i>MMP7</i> (Crom.11)	rs17098236	102387691	C	0,8836	T	0,1164	Down	N/R
	rs11568818	102401661	T	0,5411	C	0,4589	Up	N/R
	rs10895304	102389992	A	0,7362	G	0,2638	Down	N/R
	rs12285347	102396607	C	0,5206	T	0,4794	Intron	N/R
<i>ERCC1</i> (Crom.19)	rs1046282	45910672	A	0,7570	G	0,2430	3'UTR	N/R
	rs2336219	45912406	G	0,8768	A	0,1232	NS (Mi)	N/R
	rs3212980	45916441	T	0,7946	G	0,2054	Intron	N/R
	rs6509214	45981815	G	0,6850	T	0,3150	Intron	N/R
	rs10415949	45962799	A	0,7398	G	0,2602	Intron	N/R
	rs111615	45923653	A	0,5891	G	0,4109	S	N/R
	rs3212948	45924362	G	0,5959	C	0,4041	Intron	N/R
	rs3212986	45912736	C	0,7946	A	0,2054	SG	N/R
<i>ERCC2</i> (Crom.19)	rs13181	45854919	T	0,8288	G	0,1712	SG	N/R
	rs1052555	45855524	G	0,8733	A	0,1267	S	Benigno
	rs238416	45857049	T	0,5891	C	0,4109	Intron	N/R
	rs1799793	45867259	C	0,8662	T	0,1338	NS (Mi)	N/R
	rs1618536	45871606	T	0,5480	C	0,4520	Intron	N/R
	rs11878644	45876967	T	0,6081	C	0,3819	Up	N/R
<i>RBI</i> (Crom.13)	rs2227311	48987032	A	0,8681	G	0,1319	5'UTR	N/R
	rs2854344	48997693	G	0,9316	A	0,0684	Intron	N/R
<i>HNF1B</i> (Crom. 17)	rs7405776	36093022	G	0,6302	A	0,3698	Intron	N/R
	rs1016990	36088915	G	0,5206	C	0,4794	NCTrans	N/R
	rs11651052	36102381	G	0,8220	A	0,1780	Intron	N/R
	rs11658063	36103872	G	0,5891	C	0,4109	Intron	N/R
	rs11651755	36099840	T	0,5343	C	0,4657	Intron	N/R
	rs2005705	36096300	G	0,8220	A	0,1780	Intron	N/R
	rs3094509	36062299	G	0,7261	A	0,2739	Intron	N/R
	rs7501939	36101156	C	0,6576	T	0,3424	Intron	N/R
<i>NCOA3</i> (Crom. 20)	rs2076546	46268493	A	0,8699	G	0,1301	S	N/R
	rs6094752	46256424	C	0,9384	T	0,061	NS (Mi)	N/R

<i>CDHI</i> (Crom.16)	rs8056538	68802282	G	0,8014	A	0,1986	Intron	N/R
	rs8055236	83212398	G	0,7398	T	0,2602	Intron	N/R
	rs7195409	83527592	G	0,7192	A	0,2808	Intron	N/R
	rs7188750	68842895	G	0,8425	A	0,1575	NCTrans	N/R
	rs6499199	68849837	C	0,9247	T	0,0753	Intron	N/R
	rs4783689	68853671	C	0,6096	T	0,3904	Intron	N/R
	rs4782726	82701333	A	0,7535	G	0,2465	Intron	N/R
	rs4076177	68824008	A	0,6302	G	0,3698	Intron	N/R
	rs2113200	68814948	T	0,8014	A	0,1986	Intron	N/R
	rs17715799	68830511	A	0,8220	T	0,1780	Intron	N/R
	rs12919719	68822341	C	0,8357	G	0,1643	Intron	N/R

Abreviações: 3'UTR = 3 prime UTR; 5'UTR = 5 prime UTR; Crom = Cromossomo; Down = Downstream; Freq = Frequência; Intergen = Intergenica ; Mi = Missense; NCTrans = exons de transcrição não codificantes; NR = não reportado; NS = Não-sinônimo; PB = Provavelmente benigno; S = Sinônimo; SG = Stop gained; Up = Upstream.

4.2 Frequência genotípica das variantes genéticas

Foram identificadas as frequências genotípicas dos SNPs (homozigotos selvagens, heterozigotos e homozigotos para a mutação) e o *p value* do equilíbrio de Hardy-Weinberg da população de Brejo dos Santos. Os resultados da genotipagem estavam em equilíbrio de Hardy-Weinberg nos indivíduos para todos os SNPs investigados ($p > 0,05$) (APÊNDICE B). De acordo com a lei de Hardy-Weinberg, princípio fundamental na genética de populações, na ausência de fatores evolutivos (deriva genética, migração, seleção), as frequências dos genótipos e alelos permanecerão inalteradas ao longo das gerações. Nesse caso, diz-se que a população está em equilíbrio de Hardy-Weinberg (HWE) (GRAFFELMAN; WEIR et al., 2016).

4.3 Comparação das variantes de Brejo dos Santos e São Paulo

Das 90 variantes genéticas identificadas na população de longevos de Brejo dos Santos apenas 55 estavam depositadas no Banco de dados ABraOM, referentes à população de longevos de São Paulo. A figura 1 mostra a frequência do alelo menor das variantes genéticas nas duas populações.

A população da Paraíba apresentou o alelo menor mais frequente para 15 (27,27%) variantes genéticas (rs1799950, rs2048718, rs144848, rs4987047, rs2287497,

rs12951053, rs3862792, rs17655, rs2227869, rs2296148, rs3025039, rs11568818, rs2227311, rs2076546 e rs6094752). Enquanto a população de idosos de São Paulo apresentou o alelo menor mais frequente para 33 (60%) variantes genéticas (rs4986852, rs1799949, rs4986764, rs4986765, rs799923, rs206119, rs1042522, rs2909430, rs1047972, rs1801270, rs2395655, rs1799782, rs25487, rs25486, rs762507, rs4150386, rs4150360, rs4150393, rs25648, rs10434, rs3025040, rs1046282, rs2336219, rs3212980, rs11615, rs3212948, rs3212986, rs13181, rs1052555, rs1799793, rs1618536, rs7188750 e rs6499199) (Figura 1). Os idosos longevos da Paraíba e São Paulo apresentaram a mesma frequência do alelo menor para a variante rs2307177 e frequências próximas para os SNPs rs16942, rs4987117, rs4987943, rs3092856, rs25489 e rs11651755 nas duas populações.

Figura 1- Frequência alélica menor (MAF) das variantes genéticas nas populações da Paraíba e São Paulo.

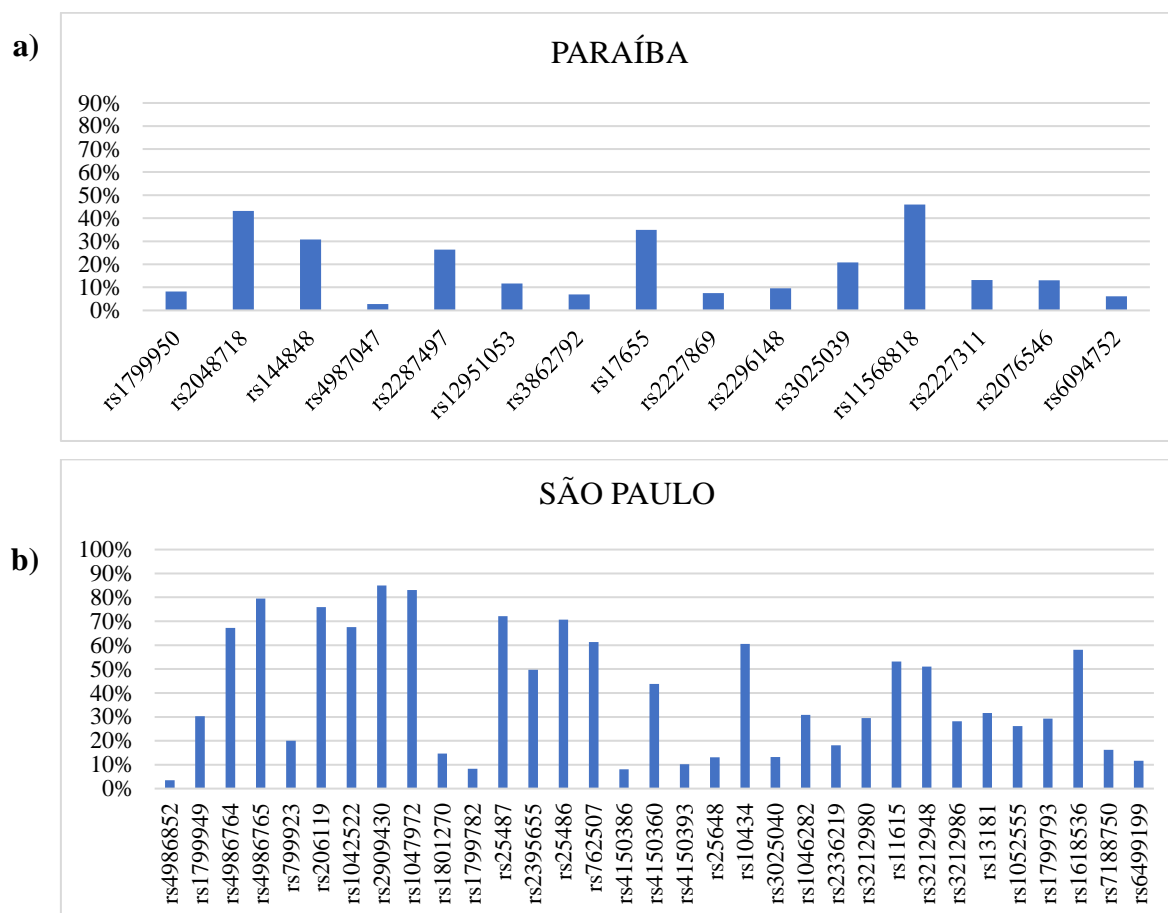


Figura 1- a) 15 SNPs que tiveram o alelo menor mais frequente na população da Paraíba, **b)** 33 SNPs que tiveram o alelo menor mais frequente na população de São Paulo.

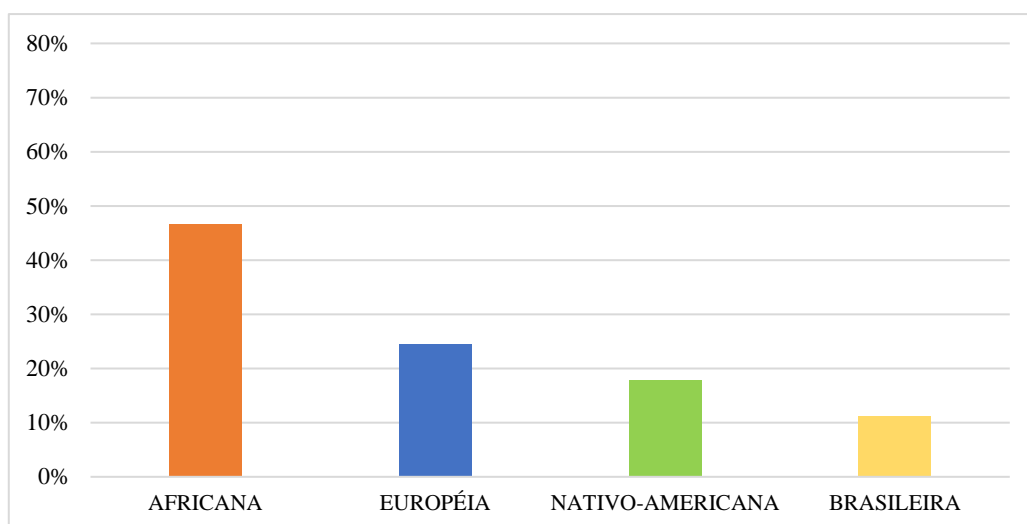
4.4 Frequência alélica menor (MAF) das variantes genéticas nas populações mundiais

A frequência alélica menor (MAF) das variantes genéticas nas principais populações do 1000 Genomas (européia, africana e nativo-americano miscigenado), foi fornecida através do “SNP reference report” (database dbSNP) (APÊNDICE C).

No conjunto de dados do projeto de 1000 genomas, 42 SNPs (46,66%) tiveram o alelo menor mais frequente na população africana, 22 SNPs (24,44%) tiveram o alelo menor mais frequente na população européia, 16 SNPs (17,77%) apresentaram o alelo menor mais frequente na população de origem nativa americana e 10 SNPs (11,11%) apresentaram o alelo menor mais frequente na população brasileira (Figura 2).

Na amostra brasileira, a frequência alélica menor (MAF) variou entre 0,0068 e 0,4794 e foi em média 0,2774 (DP = 0,138), enquanto a MAF nos dados obtidos no projeto do genoma 1000 foi de 0,3696 (DP = 0,3052) em populações de ascendência africana, 0,3062 (DP = 0,2277) em populações européias e 0,3127 (DP = 0,2431) em nativo-americanos (APÊNDICE D).

Figura 2 - Porcentagem de SNPs que apresentaram o alelo menor mais frequente nas populações do 1000 Genomas e população brasileira (Brejo dos Santos).



4.5 Inferência de Ancestralidade Local

A partir da inferência de ancestralidade local foi determinada a origem dos 90 SNPs estudados para a população de Brejo dos Santos. Ao final foram identificadas 43 variantes genéticas (47,25%) de ancestralidade européia, 45 variantes (49,45%) de

origem europeia e africana e 3 variantes genéticas (3,29%) apresentaram origem europeia e nativo-americano (Figura 3).

Figura 3- Frequência e origem do alelo maior e alelo menor dos 90 SNPs identificados na população de idosos de Brejo dos Santos.

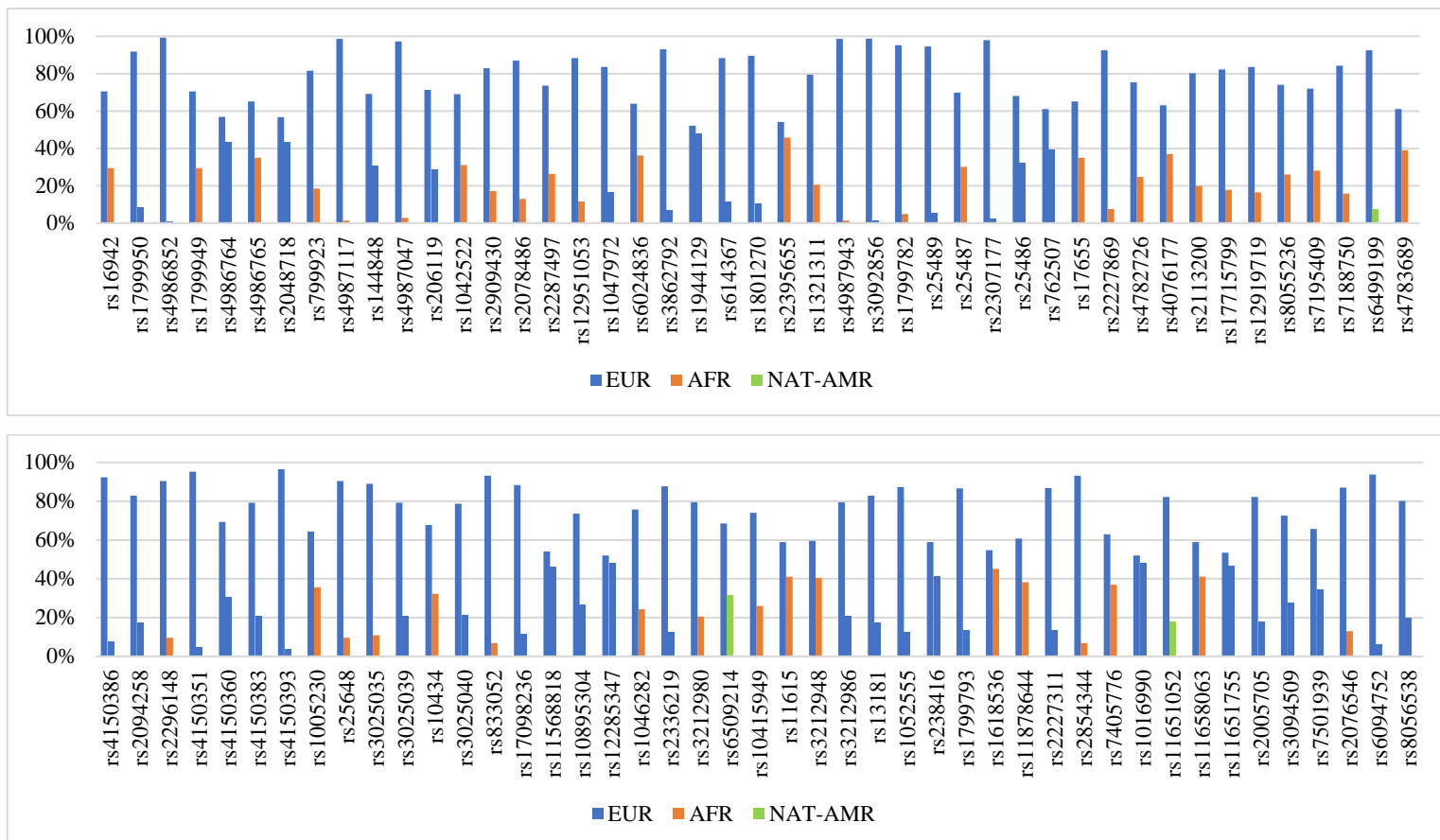


Figura 3 - A figura mostra a frequência e origem do alelo maior (primeira coluna) e alelo menor (segunda coluna) na população, sendo europeu (azul), africano (laranja) e nativo-americano (verde).

4.6 Haplótipos

A partir da análise dos resultados pelo software *Haploview*, foram identificados na população 48 haplótipos, nos seguintes cromossomos estudados 6, 11, 13, 16, 17, 19 e 20 (Figura 4). Dos 90 SNPs analisados, 35 estavam em desequilíbrio de ligação, formando os blocos de haplótipos. Os SNPs que se apresentaram dentro do bloco de LD herdaram juntos e os SNPs que ficaram fora do bloco de LD não herdaram juntos e possivelmente de forma independente. Além disso, os haplótipos frequentes na população

apresentaram frequência >20%, haplótipos considerados raros uma frequência <5% e haplótipos com frequência entre 5-20% considerados comuns.

A figura 4a mostra o cromossomo 6, em que a análise de haplótipos identificou dois blocos de ligação. No bloco 1, os SNPs rs1005230 e rs25648 apresentaram um alto desequilíbrio de ligação ($D' = 0,97$) e pertencem a um bloco de 2kb. Os três haplótipos observados nesse bloco foram CC (69,2%), TC (19,4%) e TT (11,3%). No bloco 2, as variantes rs3025039, rs3025040 e rs10134 também apresentaram alto desequilíbrio de ligação e pertencem a um bloco de 0kb. Os haplótipos desse bloco mostram as seguintes frequências de haplótipos CCG (50,1%), CCA (29,1%), TTG (18,1%) e CTG (2,3%).

O cromossomo 11 apresentou somente um bloco de ligação (Figura 4b), em que as variantes rs12285347 e rs11568818 apresentaram um forte desequilíbrio de ligação ($D' = 1$) e pertencem a um bloco de 5kb. Os haplótipos à direita apresentaram as seguintes frequências TT (53,8%), CC (43,6%) e CT (2,6%).

A análise de haplótipos identificou no cromossomo 13 apenas um bloco de ligação (Figura 4c) no qual as variantes rs4150351 e rs4150360 também apresentaram alto desequilíbrio de ligação ($D' = 1$) e pertencem a um bloco de 1kb. Os haplótipos observados neste bloco foram AC (68,3%), AT (26,2%) e CT (5,5%).

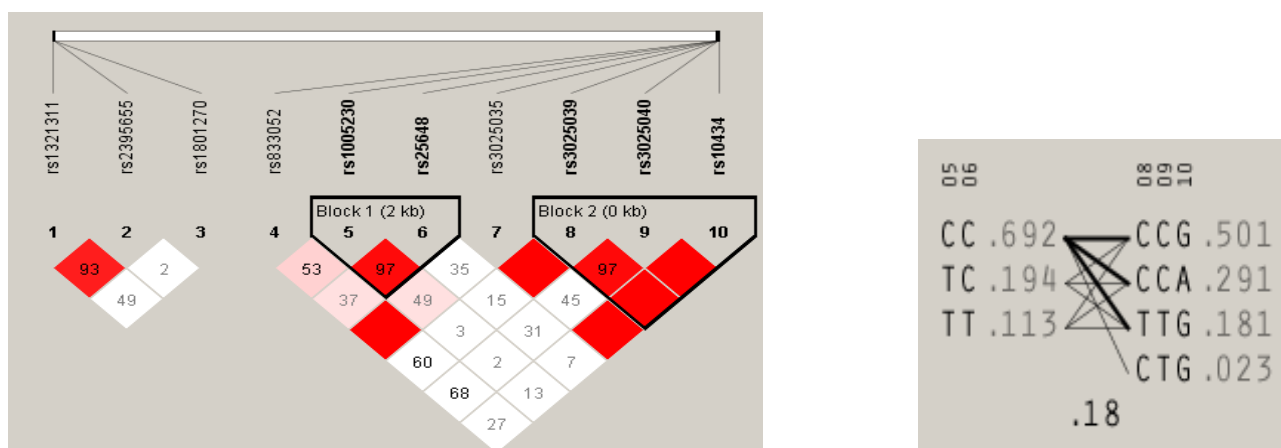
No cromossomo 16, a análise de haplótipos identificou três blocos de ligação (Figura 4d). No bloco 1 as variantes rs8056538 e rs2113200 estão em alto desequilíbrio de ligação ($D' = 1$) e pertencem a um bloco de 12kb. Os haplótipos identificados neste bloco foram GT (73,0%), AA (24,8%) e AT (2,1%). No bloco 2 as variantes rs12919719 e rs17715799 estão em desequilíbrio de ligação alto ($D' = 0,97$) e pertencem a um bloco de 8kb, apresentando os seguintes haplótipos CA (78,8%), GT (18,4%) e CT (2,4%). No bloco 3 as variantes rs7188750 e rs4783689 também estão em desequilíbrio de ligação alto ($D' = 0,96$) e pertencem a um bloco de 10kb. Os haplótipos observados neste bloco foram GC (50,1%), GT (27,5%) e AC (22,2%).

O cromossomo 17 apresentou a maior quantidade de blocos com quatro presentes (Figura 4e). As variantes rs12951053, rs2909430, rs1042522 (bloco 1- 2kb) rs7501939, rs11651052, rs11658063 (bloco 2 - 2kb) rs16942, rs1799949 (bloco 3 - 1kb) e rs4986764 e rs4986765 (bloco 4 - 0kb) apresentaram um alto desequilíbrio de ligação. O bloco 1 apresentou os seguintes haplótipos ATC (58,4%), CTG (15,0%), ACG (14,6%) e ATG (11,0%), o bloco 2 CGG (49,3%), CAG (10,8%), TAC (31,5%) e TAG (6,8%), o bloco 3 TG (69,2%), CG (2,7%) e CA (27,9%) e o bloco 4 GC (69,5%), AC (11,7%) e AT (18,8%).

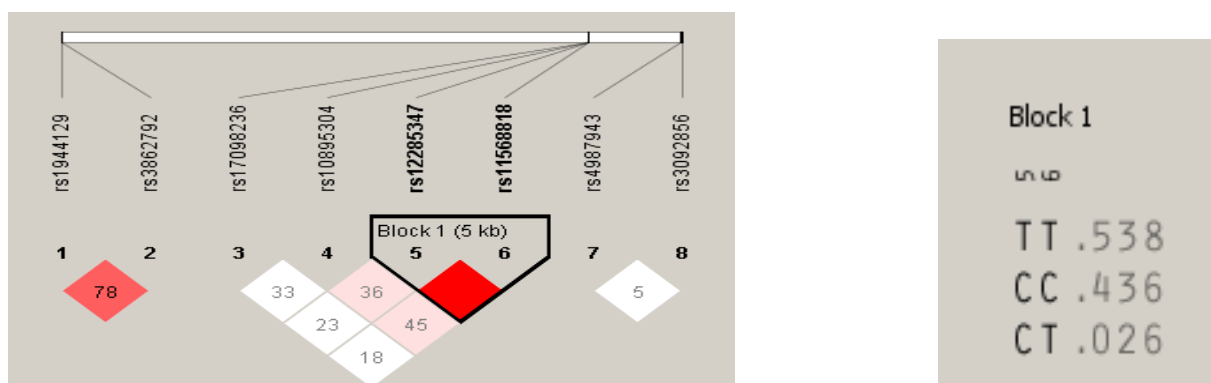
O cromossomo 19 apresentou três blocos de ligação (figura 4f). No bloco 1 as variantes rs25487, rs25486, rs1799782 e rs762607 estão em alto desequilíbrio de ligação e pertencem a um bloco de 2 kb, apresentando os seguintes haplótipos CTGC (33,5%), TCGC (28,0%), CTGT (27,6%), CTAC (8,2%) e CCGC (2,1%). No bloco 2 as variantes rs13181e rs1052555 também estão em alto desequilíbrio de ligação e pertencem a um bloco de 0 kb apresentando os haplótipos TG (78,2%), GG (7,0%) e GA (14,8%). No bloco 3 as variantes rs1046282, rs2336219, rs3212986 e rs3212980 estão alto desequilíbrio de ligação e pertencem a um bloco de 5 kb. Os haplótipos observados neste bloco foram AGCT (45,3%), AACT (21,5%), GGAG (28,4%) e GGCT (3,8%). O cromossomo 20 não apresentou blocos de haplótipos.

Figura 4 - Gráfico de desequilíbrio de ligação e frequências de haplótipos.

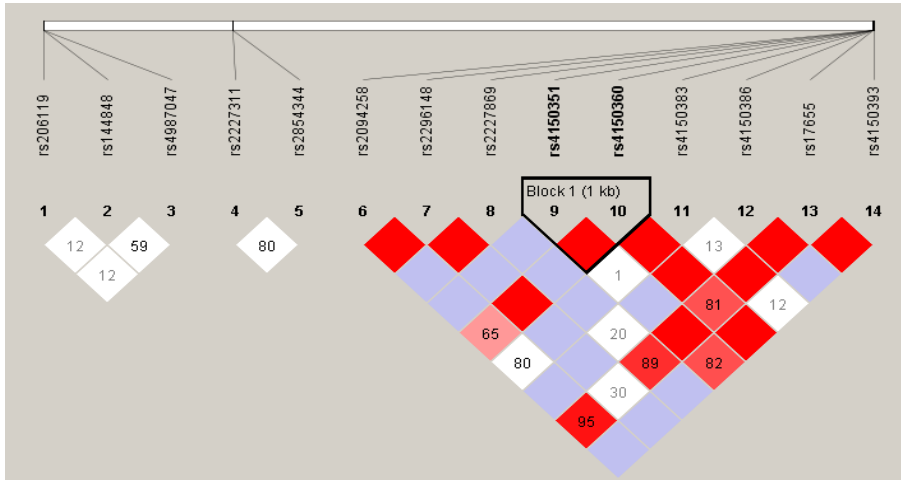
a) Cromossomo 6



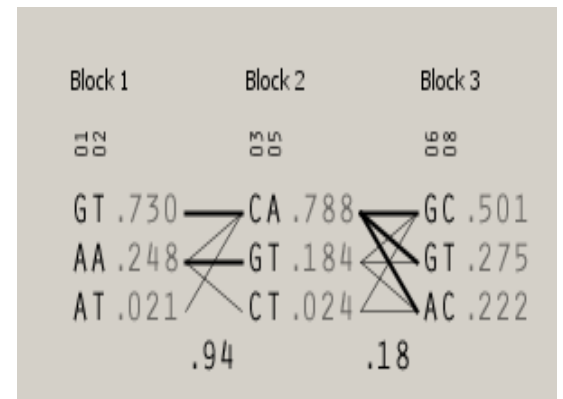
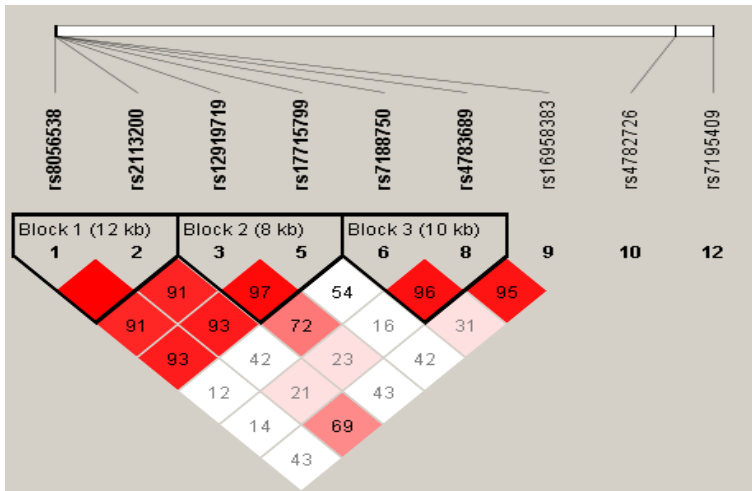
b) Cromossomo 11



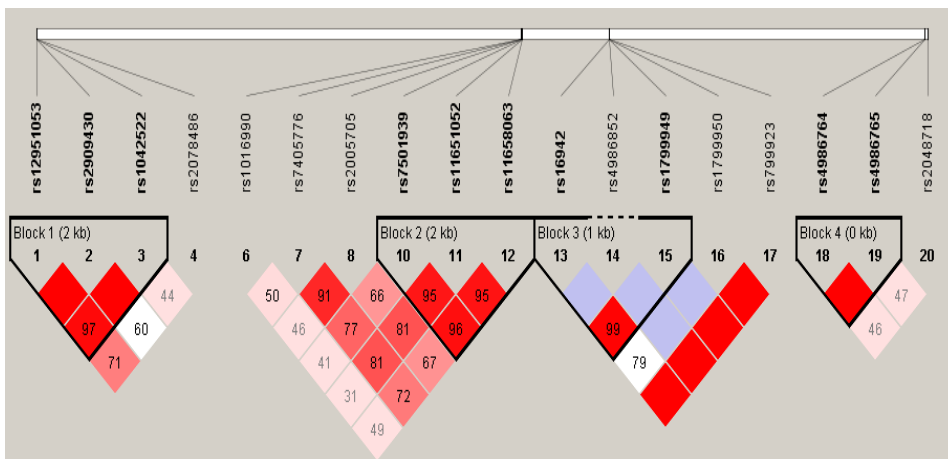
e) Cromossomo 13



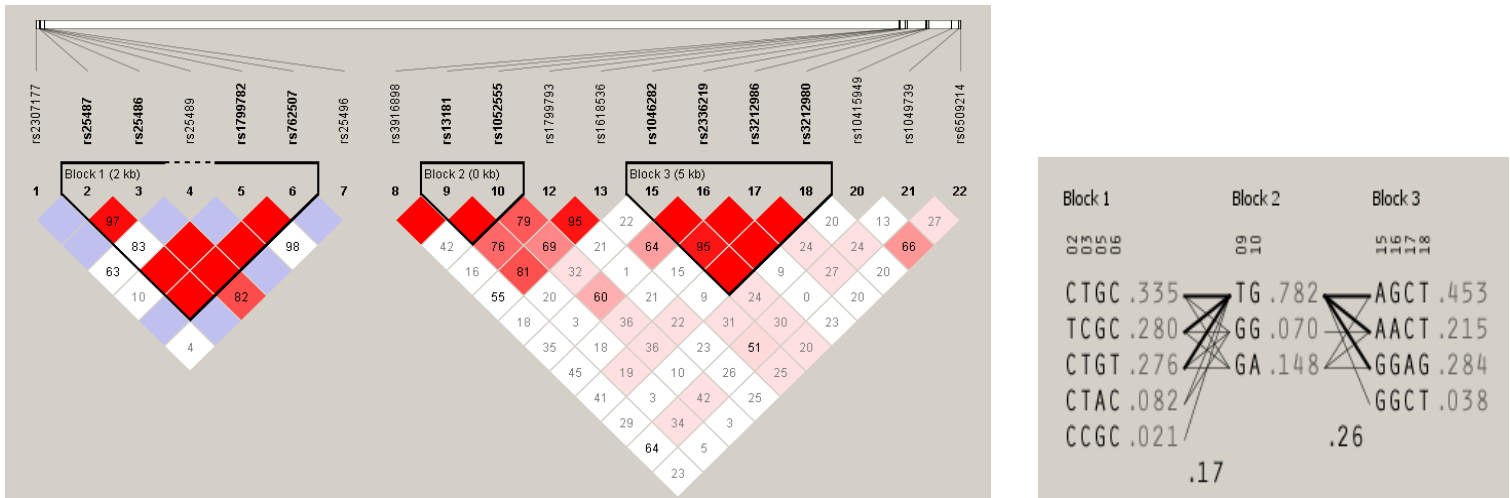
d) Cromossomo 16



e) Cromossomo 17



f) Cromossomo 19



g) Cromossomo 20

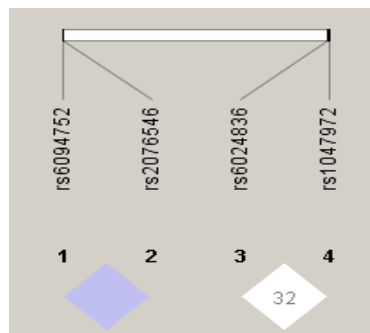


Figura 4 - Gráfico de desequilíbrio de ligação gerado pelo software Haploview 4.2 a partir dos dados dos genótipos em 73 indivíduos e frequências de haplótipos para o (a) Cromossomo 6 (b) Cromossomo 11 (c) Cromossomo 13 (d) Cromossomo 16 (e) Cromossomo 17 (f) Comossomo 19 (g) Cromossomo 20. O comprimento de cada bloco é fornecido em kilobases (kb) e o desequilíbrio de ligação aos pares (D') é fornecido para cada combinação de SNP nos cromossomos estudados. A cor vermelha mostra LD alto ($D'=1$), o que significa que 2 loci serão herdados para a próxima geração, juntamente com alta taxa de associação. A cor rosa e branca mostra LD baixo ($D'<1$). As frequências de haplótipos são exibidas à direita de cada combinação de haplótipos (cor cinza claro), enquanto o nível de recombinação é exibido acima às conexões entre dois blocos. As linhas mais grossas mostram cruzamentos mais comuns entre um bloco e outro ($>10\%$) e linhas mais finas mostrando cruzamentos menos comuns (1-10%). Essas linhas mostram as correlações entre os haplótipos. Abaixo das linhas de cruzamento está o multilocus D prime, que é uma medida do LD entre dois blocos. Quanto mais próximo de zero o valor, maior a quantidade de recombinação histórica entre os dois blocos.

Dos 48 haplótipos identificados, 26 foram considerados frequentes na população ($>20\%$) - CC (69,2%), CCG (50,1%), CCA (29,1%), TT (53,8%), CC (43,6%), AC (68,3%), AT (26,2%), GT (73,0%), AA (24,8%), CA (78,8%), GC (50,1%), GT (27,5%), AC (22,2%), ATC (58,4%), CGG (49,3%), TAC (31,5%), TG (69,2%), CA (27,9%), GC (69,5%), CTGC (33,5%), TCGC (28,0%), CTGT (27,6%), TG (78,2%), AGCT (45,3%),

AACT (21,5%) e GGAG (28,4%), 15 haplótipos considerados comuns (5-20%) - TC (19,4%), TT (11,3%), TTG (18,1%), CT (5,5%), GT (18,4%), CTG (15,0%), ACG (14,6%), ATG (11,0%), CAG (10,8%), TAG (6,8%), AC (11,7%), AT (18,8%), CTAC (8,2%), GG (7,0%), GA (14,8%) e sete haplótipos foram considerados raros na população com frequência <5%; CTG (2,3%), CT (2,6%), CG (2,7%), CCGC (2,1%), GGCT (3,8%), AT (2,1%) e CT (2,4%).

Além disso, os haplótipos GT, AA, AT (bloco 1), CA, GT, CT (bloco 2) e GC, GT, AC (bloco 3) do cromossomo 16, apresentaram o maior tamanho blocos de haplótipos, apresentando respectivamente 12kb para os haplótipos GT/AA/AT, 8kb para os haplótipos CA/GT/CT e 10kb para os haplótipos GC/GT/AC. Dessa forma, pode-se inferir que estes haplótipos são os mais compartilhados na população de Brejo dos Santos.

4.7 Origem dos haplótipos

A origem dos haplótipos foi estimada de acordo com a inferência de ancestralidade local das variantes genéticas que foram identificadas nos idosos de Brejo dos Santos no tópico 4.5. A figura 5 mostra a origem dos haplótipos por cromossomo.

No cromossomo 6, foram identificados 7 haplótipos, dois de origem europeia CC/CCA e quatro de origem europeu/africano TC/CCG/TTG/CTG e um de origem africano TT (Figura 5a). No cromossomo 11, foram identificados três haplótipos CT/CC/TT (Figura 5b) e no cromossomo 13 (Figura 5c), também três haplótipos CT/AT/AC, todos de origem europeia. No cromossomo 16, foram identificados 9 haplótipos (Figura 5d) quatro de origem europeia GT/AT/CA/GC, e cinco de origem europeu/africano AA/GT/CT/GT/AC.

No cromossomo 17, foram identificados 14 haplótipos na população (Figura 5e), destes quatro tem origem europeia ACG/CGG/TG/AT, um de origem africana CA, sete de origem europeu/africano ATC/CTG/ATG/TAG/CG/GC/AC, um europeu/nativo-americano CAG e um de origem europeu/africano/nativo-americano TAC. O cromossomo 19 apresentou 12 haplótipos (Figura 6f), desses seis de origem europeu TCGC/AGCT/AACT/TG/GG/GA e seis de origem europeu/africano CTGC/CTGT/CTAC/CCGC/GGAG/GGCT (Figura 5f).

Figura 5 - Origem ancestral dos haplótipos por cromossomo.

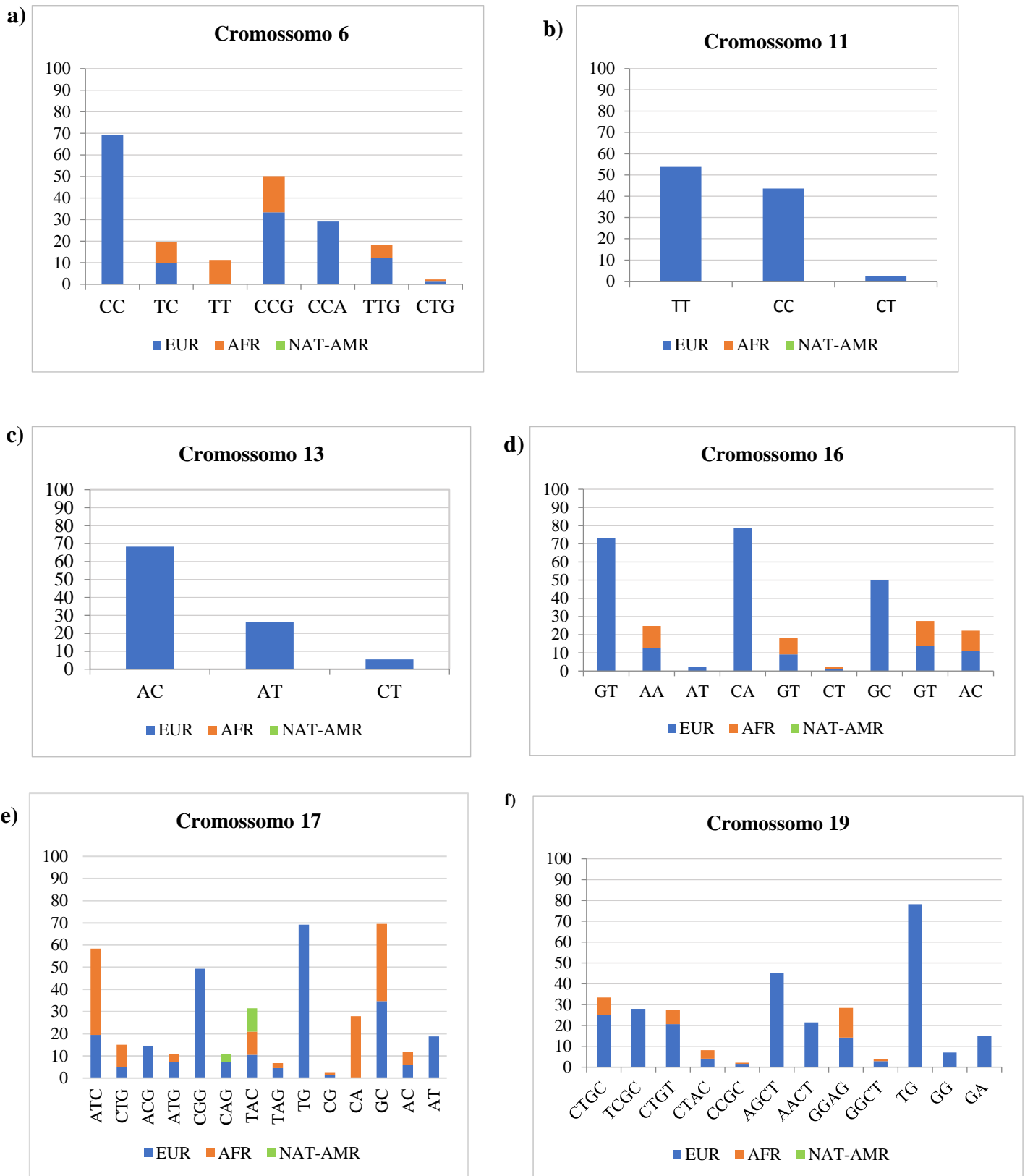
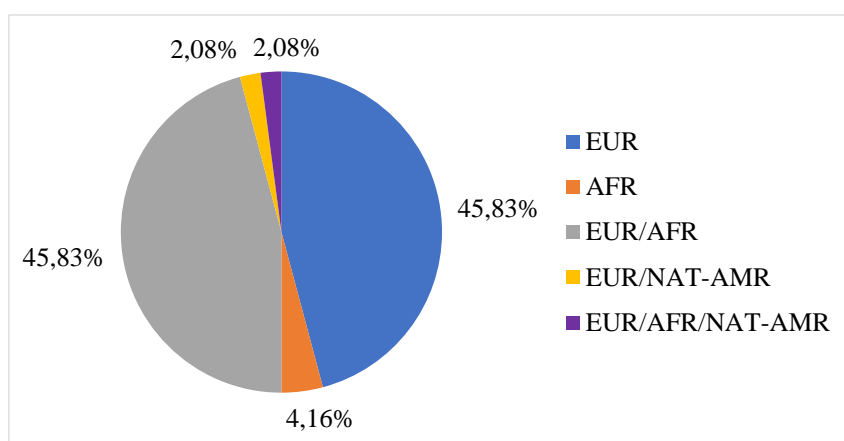


Figura 5 - EUR: Europeu, AFR: Africano, NAT-AMR: Nativo-americano miscigenado. Haplótipos de apenas uma cor, indica uma ancestralidade, haplótipos de duas cores indicam duas ancestralidades e haplótipos de três cores, três ancestralidades.

De modo geral, dos 48 haplótipos identificados na população de idosos longevos de Brejo dos Santos, 22 haplótipos (45,83%) tem ancestralidade europeia, dois haplótipos (4,16%) possui ancestralidade africana, 22 haplótipos (45,83%) tem ancestralidades europeia/africana, um haplótipo (2,08%) ancestralidade europeu/nativo-americano e um haplótipo (2,08%) apresenta três ancestralidades sendo europeu/africano/nativo-americano (Figura 6).

Figura 6 - Distribuição dos haplótipos e suas origens identificados nos idosos longevos de Brejo dos Santos.



4.8 Haplótipos e câncer

A relação entre os haplótipos e a suscetibilidade do câncer mostrou que dos 48 haplótipos identificados na população, 23 apresentaram apenas alelo de risco para o câncer: CC (69.2%), TT (11.3%), CCG (50.1%), CCA (29.1%); GT (73.0%), AA (24.8%), AT (2.1%), GT (18.4%), AC (22.2%), ATC (58.4%), CTG (15.0%), ACG (14.6%), ATG (11.0%), CAG (10.8%), CG (2.7%), CA (27.9%), CTGT (27.6%), CTAC (8.2%), GG (7.0%), GA (14.8%), AGCT (45.3%), GGAG (28.4%) e AACT (21.5%); cinco haplótipos apresentaram tanto alelo de risco quanto alelo protetor para o câncer: TTG (18.1%), CTG (2.3%), CT (2.4%), TAC (31.5%) e TAG (6.8%) e cinco haplótipos apresentaram apenas alelo protetor para o câncer: CT (5.5%); CA (78.8%), GT (27.5%), TCGC (28.0%) e GGAG (28.4%) (APÊNDICE E).

Além disso, 15 haplótipos TC (19,4%), TT (53,8%), CC (43,6%), CT (2,6%), AC (68,3%) AT (26,2%), GC (50,1%), CGG (49,3%), TG (69,2%), GC (69,5%), AC

(11,7%), AT (18,8%), CTGC (33,5%), CCGC (2,1%) e TG (78,2%) não apresentaram alelos de suscetibilidade de câncer e 12 SNPs (rs12285347, rs11568818, rs4150360, rs12951052, rs2909430, rs179949, rs4986764, rs4986765, rs25486, rs1052555, rs1046282 e rs3212980) não apresentaram alelos já conhecidos na literatura que modulam o risco de câncer e que apresentavam os mesmos alelos encontrados na população de Brejo dos Santos (APÊNDICE E).

5 DISCUSSÃO

Este é o primeiro estudo que buscou revelar a arquitetura genética para o câncer em uma população de idosos longevos que vivem no Nordeste brasileiro, onde houve intensa miscigenação entre população europeia colonizadora, nativos-americanos e africanos. Os 73 idosos não tinham relação de parentesco entre si, apresentavam grande homogeneidade no estilo de vida e não desenvolveram nenhuma forma de câncer até o momento da pesquisa.

Isto pode estar relacionado vários fatores, como hábitos alimentares saudáveis, e exercícios físicos atrelados a elevada capacidade de resiliência desses idosos enfrentando com êxito as vulnerabilidades decorrentes do processo de envelhecimento. Além disso, os fatores genéticos também podem ter influenciado, assim como a contribuição ambiental pode ter agido de forma a garantir a saúde dos longevos, apresentando dessa forma baixa predisposição ao câncer.

Em populações tri-híbridas com mesmo estilo de vida, é possível realizar estudos tipo caso-controle para melhor definir o papel de fatores genéticos e ambientais na manifestação do câncer. Um sério problema de diferenças específicas da população, em relação à etiologia, incidência e prognóstico do câncer entre indivíduos de populações humanas distintas, pode ser a discriminação entre fatores sociodemográficos, respectivamente relacionados ao estilo de vida e diferenças biológicas, com base em marcadores moleculares (ÖZDEMİR; DOTTO, 2017). Portanto, os estudos de ancestralidade que visam à detecção de novos haplótipos representando combinações de polimorfismos que influenciam a incidência do câncer de maneira diferenciada entre as populações humanas são de fundamental importância.

Na análise de variantes genéticas, os SNPs do gene *BRCA1* (rs16942, rs1799950, rs4986852, rs1799949, rs4986764, rs4986765, rs2048718 e rs799923) e do *BRCA2* (rs4987117, rs144848, rs4987047 e rs206119) foram classificadas como benignos ou provavelmente

benignos para o câncer de mama e ovário de acordo com Clinvar (Tabela 1). Da mesma forma as variantes rs1042522 (*TP53*) rs1801270 (*CDKN1A*) e todas as variantes do ATM foram classificadas como benigna ou provavelmente benignas para o câncer. As variantes que foram classificadas como benignas pelo Clinvar são interessantes para a população estudada uma vez que não apresentam histórico de câncer. Ao mesmo tempo a maioria das variantes foram consideradas de risco segundo a literatura, isto reflete que a presença de determinadas variantes genéticas pode tornar as populações mais ou menos suscetíveis ao desenvolvimento ao câncer, entretanto podem não expressar tumores uma vez que a suscetibilidade é diferente entre as populações humanas.

Essas variantes têm baixa penetrância e, apesar de terem sido associadas ao risco de câncer, elas não necessariamente causam a doença como ocorre em doenças mendelianas com alelos de alta penetrância. Um estudo com 44 centenários judeus Ashkenazi identificou mais de 130 variantes provavelmente patogênicas em seus genomas e concluíram que mutações previamente classificadas como patogênicas podem não necessariamente impedir uma vida longa (CARLSON et al., 2020). Em doenças multifatoriais ou complexas, diferentes variantes com baixa penetrância contribuem com efeito aditivo na manifestação de um determinado fenótipo patológico. Assim, uma variante pode estar associada ao fenótipo e aumentar o risco, mas ela não é determinante ou causa única do fenótipo.

Um estudo mostrou que a alta taxa e a longa história de consaguinidade, comumente observada em países em desenvolvimento, diminuem a taxa de incidência de câncer de mama ao diminuir a frequência de mutações em genes de alto nível de câncer de mama como *BRCA1* e *BRCA2* (HAMDI et al., 2018).

Na análise da frequência alélica em populações de idosos longevos de regiões geográficas diferentes, como Paraíba e São Paulo, foi possível observar que os idosos de São Paulo apresentaram o alelo menor mais frequente para 60% das variantes genéticas enquanto os idosos de Brejo dos Santos apresentaram o alelo menor mais frequente para 27,27% das variantes genéticas. Isto se deve a alguns possíveis fatores, como por exemplo, maior número de idosos presentes no banco de dados de São Paulo, além de ser uma população mais diversa e devido ao estilo de vida podem ser mais expostos a alguns fatores de risco para o câncer, ocorrendo dessa forma uma maior frequência de mutações na sua composição genética. Os idosos de Brejo dos Santos na Paraíba estão em número menor, além disso, por morarem no sertão paraibano, uma região mais isolada, possivelmente são menos expostos a fatores de risco e/ou possuem hábitos de vida mais

saudáveis. Isso reflete as diferenças genéticas entre populações de regiões geográficas diferentes, representando a elevada miscigenação da população.

Da mesma forma, foi observada diferenças na frequência do alelo menor (MAF) entre as populações mundiais do 1000 genomas (Europeu, Africano e Nativo-americano miscigenado). Dos 90 SNPs analisados, 42 variantes (46,66%) apresentaram o alelo menor mais frequente na população africana, 22 variantes (24,44%) tiveram o alelo menor mais frequente na população europeia e 16 variantes (17,77%) apresentaram o alelo menor mais frequente na população de origem nativa americana (Tabela S1). A população brasileira apresentou o alelo menor mais frequente para dez variantes genéticas (11,11%). O alelo menor do SNP rs4987047 foi encontrado exclusivamente em populações africanas (Tabela S1). Isto pode ser relacionado principalmente a diferentes fatores demográficos, ambientais e de estilo de vida, mas também resulta de diferenças na arquitetura genética de uma população para outra.

Com relação a inferência de ancestralidade local dos SNPs, a população de idosos longevos de Brejo dos Santos apresentou 43 variantes genéticas (47,25%) de origem europeia, 45 variantes (49,45%) apresentaram origem europeia e africana e 3 variantes (3,29%) apresentaram origem europeia e nativo-americana. Isto mostra uma alta contribuição europeia para a constituição genética dos indivíduos sendo superior às contribuições africana e nativo-americana miscigenada. É interessante pois há uma reduzida contribuição das populações ameríndias na população de Brejo dos Santos com apenas 3,29% das variantes. Assim, os indivíduos que vivem no sertão provavelmente têm grandes contribuições genômicas de indivíduos ibéricos. Estima-se que a população brasileira tem em média 60% de contribuição europeia, 25% da africana e 15% da população nativo-americana, entretanto essas proporções variam de acordo com a região geográfica brasileira. Por isso, a população brasileira é significativamente distinta das populações ancestrais, apresentando novos genótipos e haplótipos que podem contribuir para diferentes fenótipos, como as diversas doenças existentes (TOFOLI et al., 2016).

A análise do desequilíbrio de ligação (LD) fornece informações importantes sobre os aspectos evolutivos das populações, histórico de recombinação, além de permitir a seleção de populações e polimorfismos de nucleotídeo único (SNPs) para estudos de associação (ANDRADE, 2019). A associação entre SNP e uma doença pode ser o resultado do desequilíbrio de ligação (LD). Entretanto, a associação de alguns polimorfismos, é específica para alguns grupos populacionais. Por isso é interessante

estudos em populações miscigenadas, como a população brasileira, considerada a mais heterogênea no mundo.

A análise de haplótipos permitiu obter um perfil genético mais abrangente da população de Brejo dos Santos. 35 SNPs (rs1005230, rs25648, rs3025039, rs3025040, rs10434, rs12285347, rs11568818, rs4150351, rs4150360, rs12951053, rs2909430, rs1042522, rs7501939, rs11651052, rs11658063, rs16942, rs1799949, rs4986764, rs4986765, rs25487, rs25486, rs1799782, rs762507, rs13181, rs1052555, rs1046282, rs2336219, rs3212986, rs3212980, rs8056538, rs2113200, rs12919719, rs17715799, rs7188750 e rs4783689) se mostram em alto desequilíbrio de ligação na população.

Foram identificados na população 48 haplótipos com LD alto e que ainda não foram descritos na literatura, são novas combinações de alelos. A frequência dos haplótipos na população variou de 2.1% a 78.8%, tendo uma frequência média de 26%. Como a população longeva de Brejo dos Santos não tinham histórico de câncer, era provável que os haplótipos encontrados apresentassem mais alelos protetores do que de risco. Entretanto, 23 haplótipos apresentaram apenas alelos que aumentam o risco para o câncer, cinco haplótipos apresentaram tanto alelos que aumentam e diminuem o risco e cinco haplótipos apresentaram apenas alelo que diminui o risco para o câncer. Ademais, 15 haplótipos não apresentaram alelos que modulam o risco de câncer.

A frequência dos haplótipos portando alelos de risco variou de 2,1% a 73,0%, tendo uma frequência média de 25,80%, a frequência dos haplótipos portando alelos de risco e protetor variou de 2,3% a 31,5% tendo frequência média de 12,22% e a frequência para os haplótipos portando apenas alelos protetores variou de 5,5% a 78,8% tendo uma frequência média de 33,64%. Com isso, tem-se que os haplótipos contendo alelos protetores apresentaram a maior frequência média.

No gene *VEGF* o haplótipo CC apresentou o alelo C do rs1005230 que foi associado ao pior prognóstico de glioma em asiáticos (LINHARES et al., 2018) e o haplótipo TT apresentou o alelo T do rs25648 associado ao risco de câncer de bexiga em europeus e asiáticos (SONG et al., 2019) (APÊNDICE E).

O alelo T do rs3025039 foi associado ao risco de câncer de mama em asiáticos (SONG et al., 2019) e o alelo T do rs3025040 associado ao menor risco de carcinoma papilar de tireóide em asiáticos (LIU et al., 2017). O rs10434 apresentou o alelo G associado ao risco de câncer colorretal em asiáticos (JEON et al., 2014) e osteosarcoma (WANG, LIU et al., 2018) e o alelo A foi associado ao risco de leucemia linfocítica crônica em asiáticos (ZHU et al., 2015). Sendo assim, os haplótipos CCG e CCA

apresentaram um alelo de risco para o câncer, o haplótipo CTG apresentou um alelo de risco e um alelo protetor e o haplótipo TTG apresentou dois alelos que aumenta o risco e um alelo que diminui o risco para o câncer (APÊNDICE E).

Para o gene *MMP7* nenhum dos haplótipos apresentou alelos de risco para as variantes rs11568818 e rs12285347. Para o gene *ERCC5* apenas o haplótipo CT apresentou alelo protetor para o câncer no qual o alelo C da variante rs4150351 foi associada ao menor risco para o câncer de próstata em americanos (BARRY et al., 2012).

No gene *CDH1* as variantes rs8056538 e rs21132 foram associadas ao risco de câncer colorretal em europeus (CARVAJAL-CARMONA et al., 2010). Sendo assim, os haplótipos GT, AA e AT apresentam alelos de risco para o câncer. O alelo C do rs12919719 foi associado ao menor risco para o câncer de mama e o alelo T do rs17715799 associado ao maior risco para o câncer endometrial (GENG et al., 2018). Com isso o haplótipo CA apresenta um alelo protetor para o câncer de mama, o haplótipo GT apresenta um alelo de risco para câncer endometrial e o haplótipo CT carrega um alelo de risco e um alelo protetor para o câncer. O alelo A do rs7188750 foi associado ao menor risco para o câncer de mama em asiáticos (BEEGHLY-FADIAL et al., 2010) e o alelo T do rs4783689 associado ao menor risco para o câncer endometrial na população asiática (GENG et al., 2018). Portanto, os haplótipos GT e AC carregam pelo menos um alelo protetor para o câncer (APÊNDICE E).

No gene *TP53* os haplótipos CTG, ACG e ATG apresentaram o alelo G do rs1042522 que foi associado ao aumento no risco de câncer cervical na população subsaariana (KAMIZA et al., 2020) e carcinoma espinocelular de cabeça e pescoço em europeus (FERNÁNDEZ-MATEOS et al., 2019). Já o haplótipo ATC apresentou o alelo C do rs1042522 que foi associado ao aumento do risco de câncer retal em asiáticos (ZHANG et al., 2019) e câncer colorretal em asiáticos, europeus e americanos (ELSHAZLI et al., 2020.)

No gene *HNF1B* o alelo T do rs7501939 foi associado ao menor risco para o câncer de próstata em europeus (NIKOLIC et al., 2014), o alelo A do rs11651052 associado ao maior risco para o câncer endometrial na população europeia (PAINTER et al., 2015) e o alelo C do rs11658063 associado ao menor risco para o câncer de pulmão em afro-americanos (JONES et al., 2019). Sendo assim, o haplótipo CAG apresentou um alelo de risco, o haplótipo TAG apresentou um alelo protetor e um alelo de risco e o haplótipo TAC, um alelo de risco e um alelo protetor para o câncer (APÊNDICE E).

No gene *BRCA1* apenas os haplótipos CG e CA apresentaram o alelo C do rs16942 associado ao aumento do risco de câncer de mama em europeus e asiáticos (COX et al., 2011). No gene *XRCC1* o haplótipo TCGC apresenta o alelo T do rs25487 que foi associado ao menor risco para o câncer de bexiga em asiáticos (ZHU et al., 2016). O haplótipo CTGT apresenta o alelo T do rs762507 associado ao maior risco para o câncer esofágico em asiáticos (DAI et al., 2019) e o haplótipo CTAC apresentou o alelo A do rs1799782 associado ao maior risco para o câncer de mama na população asiática (ALIMU et al., 2018) (APÊNDICE E).

No gene *ERCC2* os haplótipos GG e GA apresentaram o alelo G do rs13181 associado ao maior risco para o câncer de laringe em europeus (FERNANDEZ-MATEOS et al., 2019), câncer de pulmão (MININA, 2019) e glioma (GAO, 2016). Além disso, o rs1052555 foi associado ao pior prognóstico para o câncer colorretal. No gene *ERCC1* o alelo A do rs2336219 foi associado a um risco maior para o câncer colorretal em asiáticos (DAI et al., 2019). O rs3212986 apresentou o alelo C associado a um maior risco para o câncer de ovário (YANG et al., 2019) e o alelo A menor risco para o câncer gástrico na população chinesa (HE et al., 2018). Assim, os haplótipos AGCT e GGAG apresentaram um alelo de risco para o câncer, o haplótipo AACT apresentou dois alelos de risco e o haplótipo GGAG apresentou um alelo protetor para o câncer (APÊNDICE E).

Os haplótipos considerados *wild-type*, apresentando apenas alelos de referência, CC (69,2%), CCA (29,1%), AC (68,3%), GT (73,0%), CA (78,8%), GC (50,1%), CCG (49,3%), TG (69,2%), TG (78,2%), AGCT (45,3%), apresentaram as maiores frequências na população. Com exceção dos haplótipos CT (2,6%) e ACG (14,6%) com frequências menores.

Nesse estudo, observou-se que os haplótipos europeus e haplótipos de origem europeu/africano apresentaram a mesma frequência na população (45,83%), com 22 haplótipos cada. Apenas dois haplótipos (4,16%) apresentaram origem africana e um haplótipo (2,08%) de origem europeu/nativo-americano e um haplótipo (2,08%) de origem europeu/africano/nativo-americano. Isto representa a alta contribuição européia e africana na população de longevos de Brejo dos Santos.

Dos 24 haplótipos que apresentam alelos de risco, dez são de origem européia, 11 de origem europeu/africano e dois de origem africano. Dos seis haplótipos que apresentam apenas alelos protetores três são de origem europeu e três de origem europeu/africano e dos 4 haplótipos que apresentam alelos de risco e protetores, três são de origem europeu/africano e um de origem europeu/africano/nativo-americano.

Os tamanhos dos blocos de haplótipos variam muito em todo o genoma, bem como entre os grupos étnicos. Os blocos do cromossomo 16, gene *CDH1*, foram os que se apresentaram em maior tamanho. Os haplótipos GT, AA, AT estão localizados em um bloco de 12 kb, os haplótipos CA, GT, e CT em um bloco de 8 kb e os haplótipos GC, GT e AC em um bloco de 10 kb. Os haplótipos GT, AT, CA, GC tem origem europeu e os haplótipos AA, GT,CT,GT e AC origem europeu/africano. Os haplótipos africanos TT e CA estão localizados em um bloco de 2 kb e 1kb respectivamente.

Isto vai de acordo com a literatura em que populações européias têm tamanhos médios de blocos de haplótipos de aproximadamente 18 kb, sendo maiores que as populações africanas. As populações de origem africana, que são evolutivamente mais antigas que as populações europeias, tendem a ter blocos menores de haplótipos - com tamanhos médios de blocos de haplótipos de aproximadamente 9 kb (SILVERMAN, 2007). Dessa forma, pode-se inferir que os haplótipos do cromossomo 16 são os mais compartilhados na população de Brejo dos Santos. Além disso, um compartilhamento grande de haplotipo pode indicar que a mutação é recente.

Quando populações previamente isoladas como a de Brejo dos Santos se reúnem e trocam material genético, ocorre a mistura genética. Essa mistura pode permitir uma rápida evolução adaptativa nas populações humanas, introduzindo novas variantes genéticas (haplótipos) com diferentes frequências. Por isso, as populações latino-americanas miscigenadas são consideradas ideal para estudar os efeitos da evolução adaptativa em humanos (NORRIS et al., 2020). Além disso, é importante estabelecer o impacto dos haplótipos na suscetibilidade a doenças, uma vez que várias gerações de recombinação geram novos padrões de ancestralidade.

Dados epidemiológicos destacam grandes disparidades raciais na incidência e sobrevivência de muitos tipos de câncer. Além de fatores genéticos, diferenças nos estilos de vida, especificamente na dieta, têm sido propostas como uma possível explicação para essa observação. As diferenças raciais em uma ampla variedade de fenótipos e suscetibilidade a doenças podem ser atribuídas em parte à diversidade genômica. Estudos comparativos mostram maior diversidade genética e níveis mais baixos de desequilíbrio de ligação nas populações africanas em relação a todas as outras populações não africanas. Portanto, a identificação de mutações causadoras de câncer é essencial para uma melhor compreensão da carcinogênese de diferentes origens genéticas (ÖZDEMIR, DOTTO; 2017).

Na última década, estudos em populações miscigenadas ganharam cada vez mais interesse na medicina e na genética de populações. Até agora, esses estudos contribuíram na compreensão sobre os padrões de variação genética ao longo da evolução humana, na demografia e dos processos adaptativos das populações humanas (GEZA et al., 2019). Permitem ainda identificar genes que conferem risco diferencial entre populações, distinguindo efeitos biológicos devidos à ancestralidade (SJAARDA, et al., 2020).

O mapeamento de populações miscigenadas levou à descoberta de muitos genes associados ao risco diferencial de doença por ancestralidade, destacando a importância de abordagens baseadas em ancestrais para estudos de associação (ZIYATDINOV et al., 2019). Devido as diferenças globais nas frequências de câncer entre as populações, se torna fundamental compreender a interação entre ancestralidade genética e influência ambiental na incidência de câncer nas populações (CORDO et al., 2020). Dessa forma, os estudos de ancestralidade são aplicados com objetivo de identificar variantes genéticas associadas a doenças que diferem em frequência entre os grupos ancestrais. Essa abordagem baseia-se no pressuposto de que a frequência dos alelos de risco varia entre as populações. Assim, diferenças no risco entre grupos ancestrais podem ser observados (SJAARDA, et al., 2020).

Uma das limitações do presente estudo foi o baixo número da amostra. Entretanto, a função potencial dos 48 novos haplótipos é desconhecida e deve ser considerada. Os longevos de Brejo dos Santos permitiram a construção de um banco de dados genético para uso em estudos adicionais de associação ao câncer relacionado à ancestralidade. Estudos do tipo GWAS e, posteriormente, de expressão gênica, se fazem necessários, a fim de verificar seus possíveis efeitos biológicos em populações miscigenadas. Análises de GWAS podem ainda detectar associações significativas devido ao LD com possíveis mutações causais que afetem características de interesse.

6 CONCLUSÃO

Este foi o primeiro estudo de investigação de variantes genéticas de suscetibilidade ao câncer e de ancestralidade realizado na população de longevos de Brejo dos Santos. Foram identificadas 90 variantes genéticas de suscetibilidade para o câncer e 48 haplótipos em alto desequilíbrio de ligação na população. Destes, 33 haplótipos foram associados a suscetibilidade de câncer. A população apresentou 22 haplótipos de origem europeia e 22 haplótipos de origem europeu/africano e pela primeira vez dois haplótipos de ascendência nativo-americana. Os resultados apresentados fornecem o primeiro

relatório sobre polimorfismos genéticos do câncer na população, sendo possível estabelecer um banco de dados de referência para orientar futuros estudos de associação em coortes de idosos longevos.

REFERÊNCIAS

1000 GENOMES PROJECT CONSORTIUM et al. An integrated map of genetic variation from 1,092 human genomes. **Nature**, v. 491, n. 7422, p. 56-65, 2012.

ALAO, J.P. The regulation of cyclin D1 degradation: roles in cancer development and the potential for therapeutic invention. **Molecular cancer**, v. 6, n. 1, p. 24, 2007.

ALIMU, N. et al. The association between XRCC1 polymorphism and laryngeal cancer susceptibility in different ethnic groups in Xinjiang, China. **International Journal of Clinical and Experimental Pathology**, v. 11, n. 9, p. 4595, 2018.

ANDRADE, A.C.B. et al. Linkage disequilibrium and haplotype block patterns in popcorn populations. **PloS one**, v. 14, n. 9, 2019.

ASHTON-PROLLA, P; SEUANEZ, H.N. The Brazilian Hereditary Cancer Network: historical aspects and challenges for clinical cancer genetics in the public health care system in Brazil. **Genet. Mol. Biol.**, Ribeirão Preto, v. 39, n. 2, p. 163-165, 2016.

AUNAN, J.R.; CHO, W.C.; SØREIDE, K. The biology of aging and cancer: a brief overview of shared and divergent molecular hallmarks. **Aging and disease**, v. 8, n. 5, p. 628, 2017.

BAO, Y. et al. Role of common ERCC1 polymorphisms in cisplatin-resistant epithelial ovarian cancer patients: A study in Chinese cohort. **International Journal of Immunogenetics**, 2020.

BARRY, K. H et al. Genetic variation in nucleotide excision repair pathway genes, pesticide exposure and prostate cancer risk. **Carcinogenesis**, v. 33, n. 2, p. 331-337, 2012.

BEEGHLY FADIEL, A. et al. Genetic polymorphisms in the MMP-7 gene and breast cancer survival. **International journal of cancer**, v. 124, n. 1, p. 208-214, 2009.

BEEGHLY-FADIEL, A. et al. Common MMP-7 polymorphisms and breast cancer susceptibility: a multistage study of association and functionality. **Cancer research**, v. 68, n. 15, p. 6453-6459, 2008.

BEEGHLY-FADIEL, Alicia et al. E-cadherin polymorphisms and breast cancer susceptibility: a report from the Shanghai Breast Cancer Study. **Breast cancer research and treatment**, v. 121, n. 2, p. 445-452, 2010.

BERNDT, S. I. et al. Large-scale fine mapping of the HNF1B locus and prostate cancer risk. **Human molecular genetics**, v. 20, n. 16, p. 3322-3329, 2011.

BERRY, J.L. et al. The RB1 Story: Characterization and Cloning of the First Tumor Suppressor Gene. **Genes**, v. 10, n. 11, p. 879, 2019.

BIAŁKOWSKA, K. et al. Association of zinc level and polymorphism in MMP-7 gene with prostate cancer in Polish population. **PloS one**, v. 13, n. 7, 2018.

BILOUS, N. et al. Clinical relevance of TP53 polymorphic genetic variations in chronic lymphocytic leukemia. **Leukemia research**, v. 58, p. 1-8, 2017.

BOLDRIN, E. et al. Association between ERCC1 rs3212986 and ERCC2/XPD rs1799793 and OS in Patients with Advanced Esophageal Cancer. **Frontiers in oncology**, v. 9, p. 85, 2019.

BOSE, S. et al. The ATM tumour suppressor gene is down regulated in EBV associated nasopharyngeal carcinoma. **The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland**, v. 217, n. 3, p. 345-352, 2009.

BRASIL. Instituto Brasileiro de geografia e Estatística - IBGE. 2017.

BURGHHAUS, S. et al. Genetic risk factors for ovarian cancer and their role for endometriosis risk. **Gynecologic oncology**, v. 145, n. 1, p. 142-147, 2017.

BURWINKEL, B. et al. Association of NCOA3 polymorphisms with breast cancer risk. **Clinical cancer research**, v. 11, n. 6, p. 2169-2174, 2005.

CARLSON, P. et al. Prevalence of clinically actionable disease variants in exceptionally long-lived families. **BMC Medical Genomics**, v. 13, n. 1, p. 1-7, 2020.

CARVAJAL-CARMONA, L. G. et al. Fine-mapping of colorectal cancer susceptibility loci at 8q23. 3, 16q22. 1 and 19q13. 11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. **Human molecular genetics**, v. 20, n. 14, p. 2879-2888, 2011.

CHANG, B. L. et al. Validation of genome-wide prostate cancer associations in men of African descent. **Cancer Epidemiology and Prevention Biomarkers**, v. 20, n. 1, p. 23-32, 2011.

CHANG, J. S. et al. Base excision repair genes and risk of lung cancer among San Francisco Bay Area Latinos and African-Americans. **Carcinogenesis**, v. 30, n. 1, p. 78-87, 2009.

CHANG, J. S. et al. Nucleotide excision repair genes and risk of lung cancer among San Francisco Bay Area Latinos and African Americans. **International journal of cancer**, v. 123, n. 9, p. 2095-2104, 2008.

CHEN, J.S. et al. Identification of novel markers for monitoring minimal residual disease in acute lymphoblastic leukemia. **Blood**, v. 97, n. 7, p. 2115-2120, 2001.

CHEN, J; WANG, H; LI, Z. Association between polymorphisms of X-ray repair cross complementing group 1 gene and pancreatic cancer risk: a systematic review with meta-analysis. **Pathology & Oncology Research**, v. 25, n. 3, p. 897-904, 2019.

CONCANNON, P. et al. Variants in the ATM gene associated with a reduced risk of contralateral breast cancer. **Cancer research**, v. 68, n. 16, p. 6486-6491, 2008.

CORDO, P.C et al. Geographic assessment of cancer genome profiling studies. **bioRxiv**, p. 827683, 2020.

COX, D. G. et al. Common variants of the BRCA1 wild-type allele modify the risk of breast cancer in BRCA1 mutation carriers. **Human molecular genetics**, v. 20, n. 23, p. 4732-4747, 2011.

DAI, L. et al. Association between intronic polymorphisms of XRCC1, ERCC2 and LIG1 genes and risk of esophageal squamous cell carcinoma in a Chinese Han population. **International journal of Clinical and Experimental Medicine**, v. 12, n. 3, p. 2710-2719, 2019.

DAI, Q. et al. XRCC1 and ERCC1 polymorphisms are related to susceptibility and survival of colorectal cancer in the Chinese population. **Mutagenesis**, v. 30, n. 3, p. 441-449, 2015.

DE FARIAS, A.A. et al. Origin and age of the causative mutations in KLC2, IMPA1, MED25 and WNT7A unravelled through Brazilian admixed populations. **Scientific reports**, v. 8, n. 1, p. 1-8, 2018.

DENG, N. et al. Single nucleotide polymorphisms and cancer susceptibility. **Oncotarget**, v. 8, n. 66, p. 110635, 2017.

DI LIU, Y. Z. et al. Four common polymorphisms of BRIP1 (rs2048718, rs4988344, rs4986764, and rs6504074) and cancer risk: evidence from 13,716 cancer patients and 15,590 cancer-free controls. **Ageing (Albany NY)**, v. 10, n. 2, p. 266, 2018.

DOHERTY, J.A. et al. Polymorphisms in nucleotide excision repair genes and endometrial cancer risk. **Cancer Epidemiology and Prevention Biomarkers**, v. 20, n. 9, p. 1873-1882, 2011.

DUFLOTH, R.M. et al. Analysis of BRCA1 and BRCA2 mutations in Brazilian breast cancer patients with positive family history. **Sao Paulo Med. J.**, v. 123, n. 4, p. 192-197, 2005.

DUTTO, I. et al. Biology of the cell cycle inhibitor p21 CDKN1A: molecular mechanisms and relevance in chemical toxicology. **Archives of toxicology**, v. 89, n. 2, p. 155-178, 2015.

ELSHAZLI, Rami M. et al. Genetic polymorphisms of TP53 (rs1042522) and MDM2 (rs2279744) and colorectal cancer risk: An updated meta-analysis based on 59 case-control studies. **Gene**, v. 734, p. 144391, 2020.

ERICHSEN, H.C.; CHANOCK, S.J. SNPs in cancer research and treatment. **British journal of cancer**, v. 90, n. 4, p. 747, 2004.

FÁ, M. de et al. Envelhecimento bem-sucedido em idosos longevos: uma revisão integrativa. **Geriatrics, Gerontology and Aging**, v. 9, n. 2, p. 66-70, 2015.

FERNÁNDEZ-MATEOS, J. et al. Genetic susceptibility in head and neck squamous cell carcinoma in a Spanish population. **Cancers**, v. 11, n. 4, p. 493, 2019.

FETT-CONTE, A.C.; SALLES, A.B.C.F. A importância do gene p53 na carcinogênese humana. **Revista Brasileira de Hematologia e Hemoterapia**, v. 24, n. 2, p. 85-89, 2002.

FURRIOL, J. et al. Associations between VEGF polymorphisms and clinical outcome in breast cancer. 2015.

GAO, X. et al. ERCC2 rs13181 polymorphism association with glioma susceptibility in a Chinese population. **Genet Mol Res**, v. 15, n. 2, 2016.

GENG, T.T et al. Association of colorectal cancer susceptibility variants with esophageal cancer in a Chinese population. **World Journal of Gastroenterology: WJG**, v. 21, n. 22, p. 6898, 2015.

GENG, Y.H. et al. Genetic polymorphisms in CDH1 are associated with endometrial carcinoma susceptibility among Chinese Han women. **Oncology letters**, v. 16, n. 5, p. 6868-6878, 2018.

GEZA, E. et al. A comprehensive survey of models for dissecting local ancestry deconvolution in human genome. **Briefings in bioinformatics**, v. 20, n. 5, p. 1709-1724, 2019.

GHONCHEH, M; POURNAMDAR, Z; SALEHINIYA, H. Incidence and mortality and epidemiology of breast cancer in the world. **Asian Pacific Journal of Cancer Prevention**, v. 17, n. S3, p. 43-46, 2016.

GÓMEZ-DÍAZ, B. et al. Analysis of ERCC1 and ERCC2 gene variants in osteosarcoma, colorectal and breast cancer. **Oncology letters**, v. 9, n. 4, p. 1657-1661, 2015.

GONZÁLEZ, C. A. et al. Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). **International journal of cancer**, v. 107, n. 4, p. 629-634, 2003.

GOODE, E. L. Linkage disequilibrium. **Encyclopedia of Cancer**, p. 2043-2048, 2011.

GUO, M. et al. Increased AURKA promotes cell proliferation and predicts poor prognosis in bladder cancer. **BMC systems biology**, v. 12, n. 7, p. 11-17, 2018.

GRAFFELMAN, J; WEIR, B.S. Testing for Hardy–Weinberg equilibrium at biallelic genetic markers on the X chromosome. **Heredity**, v. 116, n. 6, p. 558-568, 2016.

HAMDI, Y. et al. A genome wide SNP genotyping study in the Tunisian population: specific reporting on a subset of common breast cancer risk loci. **BMC cancer**, v. 18, n. 1, p. 1-14, 2018.

HAN, L. A case-control study on the genetic risks for development of EGFR mutant and wild-type lung adenocarcinoma in never-smoking Chinese population. **HKU Theses Online (HKUTO)**, 2017.

HAN, M.R. et al. Evaluating genetic variants associated with breast cancer risk in high and moderate-penetrance genes in Asians. **Carcinogenesis**, v. 38, n. 5, p. 511-518, 2017.

HAO, W; STOREY, J.D. Extending tests of Hardy–Weinberg equilibrium to structured populations. **Genetics**, v. 213, n. 3, p. 759-770, 2019.

HARTMAIER, R. J. et al. Nuclear receptor coregulator SNP discovery and impact on breast cancer risk. **BMC cancer**, v. 9, n. 1, p. 438, 2009.

HARTMAN, M.L.; CZYZ, M. MITF in melanoma: mechanisms behind its expression and activity. **Cellular and molecular life sciences**, v. 72, n. 7, p. 1249-1260, 2015.

HE, Jing et al. Genetic variants in the nucleotide excision repair pathway genes and gastric cancer susceptibility in a southern Chinese population. **Cancer management and research**, v. 10, p. 765, 2018.

HE, M. G. et al. Association between ERCC1 and ERCC2 gene polymorphisms and susceptibility to pancreatic cancer. **Genet Mol Res**, v. 15, n. 1, 2016.

HOFFMANN, T.J. et al. Genome-wide association study of prostate-specific antigen levels identifies novel loci independent of prostate cancer. **Nature communications**, v. 8, n. 1, p. 1-11, 2017.

HOOKER, S. et al. NAT2 and NER genetic variants and sporadic prostate cancer susceptibility in African Americans. **Prostate cancer and prostatic diseases**, v. 11, n. 4, p. 349-356, 2008.

HUANG Y. L et al. The Association Between Polymorphism of ERCC1 Gene and Hepatocellular Carcinoma. *Labeled Immunoassays and Clinical Medicine*. v.25, n. 10, p. 1469-1475, 2018.

HUSSAIN, S. K. et al. Genetic variation in immune regulation and DNA repair pathways and stomach cancer in China. **Cancer Epidemiology and Prevention Biomarkers**, v. 18, n. 8, p. 2304-2309, 2009.

JOHNATTY, S. E. et al. Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and TERT, a cancer susceptibility “hot-spot”. **PLoS Genet**, v. 6, n. 7, p. e1001016, 2010.

JOHNSON, N. et al. Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility. **Human molecular genetics**, v. 16, n. 9, p. 1051-1057, 2007.

JONES, C. C. et al. Cross-cancer pleiotropic associations with lung cancer risk in African Americans. **Cancer Epidemiology and Prevention Biomarkers**, v. 28, n. 4, p. 715-723, 2019.

KAGUE, E. et al. Methylation status of CDH1 gene in samples of gastric mucous from brazilian patients with chronic gastritis infected by *Helicobacter pylori*. **Arq. Gastroenterol.**, v. 47, n. 1, p. 7-12, 2010.

KAMIZA, A. B et al. Association of TP53 rs1042522 with cervical cancer in the sub-Saharan African population: a meta-analysis. **Tropical Medicine & International Health**, 2020.

KIDO, T. et al. Are minor alleles more likely to be risk alleles?. **BMC medical genomics**, v. 11, n. 1, p. 3, 2018.

LAMBRECHTS, D. et al. 11q13 is a susceptibility locus for hormone receptor positive breast cancer. **Human mutation**, v. 33, n. 7, p. 1123-1132, 2012.

LEBRÃO, M.L. et al. 10 Anos do Estudo SABE: antecedentes, metodologia e organização do estudo. **Revista Brasileira de Epidemiologia**, v. 21, p. e180002, 2019.

LEE, S.Y. et al. Functional intronic ERCC1 polymorphism from regulomeDB can predict survival in lung cancer after surgery. **Oncotarget**, v. 6, n. 27, p. 24522, 2015.

LEMES, M. R; ALVES, L.C.C.B; YAMAGUCHI, M.U. Nível de resiliência em idosos segundo a escala de Connor-Davidson: uma revisão sistemática. **Rev. bras. geriatr. gerontol.**, v. 22, n. 3, e180209, 2019.

LESUEUR, F. et al. Single-nucleotide polymorphisms in the RB1 gene and association with breast cancer in the British population. **British journal of cancer**, v. 94, n. 12, p. 1921-1926, 2006.

LEWANDOWSKA, Anna Maria et al. Environmental risk factors for cancer-review paper. **Annals of agricultural and environmental medicine: AAEM**, v. 26, n. 1, p. 1-7, 2018.

LI, B. et al. ERCC1 rs11615 polymorphism increases susceptibility to breast cancer: a meta-analysis of 4547 individuals. **Bioscience reports**, v. 38, n. 3, 2018.

LI, Q. et al. Association between the BRCA2 rs144848 polymorphism and cancer susceptibility: a meta-analysis. **Oncotarget**, v. 8, n. 24, p. 39818, 2017.

LI, Y. K. et al. Nucleotide excision repair pathway gene polymorphisms are associated with risk and prognosis of colorectal cancer. **World Journal of Gastroenterology**, v. 26, n. 3, p. 307, 2020.

LI, Z. et al. Amplifications of NCOA3 gene in colorectal cancers in a Chinese population. **World Journal of Gastroenterology: WJG**, v. 18, n. 8, p. 855, 2012.

LIMA, S. et al. A population-based study of inter-generational attitudes towards consanguineous marriages in north-eastern Brazil. **Journal of biosocial science**, v. 51, n. 5, p. 683-697, 2019.

LIN, Y. et al. Polymorphism rs144848 in BRCA2 may reduce lung cancer risk in women: a case-control study in southeast China. **Tumori Journal**, v. 102, n. 2, p. 150-155, 2016.

LINHARES, P. et al. Genetic variants of vascular endothelial growth factor predict risk and survival of gliomas. **Tumor Biology**, v. 40, n. 3, p. 1010428318766273, 2018.

LIU, R. et al. Association between single nucleotide variants of vascular endothelial growth factor A and the risk of thyroid carcinoma and nodular goiter in a Han Chinese population. **Oncotarget**, v. 8, n. 9, p. 15838, 2017.

LOIZIDOU, M. A. et al. Genetic polymorphisms in the DNA repair genes XRCC1, XRCC2 and XRCC3 and risk of breast cancer in Cyprus. **Breast cancer research and treatment**, v. 112, n. 3, p. 575-579, 2008.

LOURENCO, T.M. et al. Capacidade funcional no idoso longo vivo: uma revisão integrativa. **Rev. Gaúcha Enferm.**, Porto Alegre, v. 33, n. 2, p. 176-185, 2012.

MA, B; WU, J; MA, B. Association between single nucleotide polymorphism of BRCA1-interacting protein C-terminal helicase 1 and early-onset breast cancer in Uygur and Han women in Xinjiang of China. **Biotechnology & Biotechnological Equipment**, v. 33, n. 1, p. 1750-1756, 2019.

MA, H. et al. Polymorphisms of XPG/ERCC5 and risk of squamous cell carcinoma of the head and neck. **Pharmacogenetics and genomics**, v. 22, n. 1, p. 50, 2012.

MAHJABEEN, I. et al. Genetic variations in XRCC1 gene in sporadic head and neck cancer (HNC) patients. **Pathology & Oncology Research**, v. 19, n. 2, p. 183-188, 2013.

MAPLES, B.K. et al. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. **The American Journal of Human Genetics**, v. 93, n. 2, p. 278-288, 2013.

MARTINEZ, C.A.R et al. Avaliação da expressão tecidual do gene de reparo MLH1 e dos níveis de dano oxidativo ao DNA em doentes com câncer colorretal. **Rev bras. coloproctol.**, v. 29, n. 3, p. 303-313, 2009.

MEDREK, K. et al. Association of common WRAP 53 variant with ovarian cancer risk in the Polish population. **Molecular biology reports**, v. 40, n. 3, p. 2145-2147, 2013.

MESIC, A. et al. Characterization and risk association of polymorphisms in Aurora kinases A, B and C with genetic susceptibility to gastric cancer development. **BMC cancer**, v. 19, n. 1, p. 919, 2019.

MININA, V. I. et al. Polymorphisms in DNA repair genes in lung cancer patients living in a coal-mining region. **European Journal of Cancer Prevention**, v. 28, n. 6, p. 522-528, 2019.

MUTSCHALL, L. et al. Análise da relação entre a expressão da proteína VEGF e o estadiamento do câncer colorretal. **Rev bras. colo-proctol.**, v. 29, n. 1, p. 15-22, 2009.

NASLAVSKY, M.S. et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. **Human Mutation**, v. 38, n. 7, p. 751-763, 2017.

NIKOLIĆ, Z.Z. et al. Assessment of Association between Common Variants at 17q12 and Prostate Cancer Risk—Evidence from Serbian Population and Meta-Analysis. **Clinical and translational science**, v. 7, n. 4, p. 307-313, 2014.

NORRIS, E.T. et al. Admixture-enabled selection for rapid adaptive evolution in the Americas. **Genome Biology**, v. 21, n. 1, p. 1-12, 2020.

OH, J.J. et al. Genome-wide detection of allelic genetic variation to predict advanced-stage prostate cancer after radical prostatectomy using an exome SNP chip. In: **Urologic Oncology: Seminars and Original Investigations**. Elsevier, 2015. p. 385. e7-385. e13.

ÖZDEMİR, B.C.; DOTTO, G.P. Racial differences in cancer susceptibility and survival: more than the color of the skin?. **Trends in cancer**, v. 3, n. 3, p. 181-197, 2017.

PAINTER, J. N. et al. Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. **Human molecular genetics**, v. 24, n. 5, p. 1478-1492, 2015.

PARK, J.Y. et al. Polymorphism of the DNA repair gene XRCC1 and risk of primary lung cancer. **Cancer Epidemiology and Prevention Biomarkers**, v. 11, n. 1, p. 23-27, 2002.

PECORINO, L. Molecular biology of cancer: mechanisms, targets, and therapeutics. **Oxford University Press**, USA, 2016.

QIN, H. D. et al. Comprehensive pathway-based association study of DNA repair gene variants and the risk of nasopharyngeal carcinoma. **Cancer research**, v. 71, n. 8, p. 3000-3008, 2011.

QING, T. et al. Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. **Nature communications**, v. 11, n. 1, p. 1-8, 2020.

RASHKIN, S. R. et al. Pan-Cancer Study Detects Novel Genetic Risk Variants and Shared Genetic Basis in Two Large Cohorts. **bioRxiv**, p. 635367, 2019.

REISBERG, S. et al. Comparison of variation in frequency for SNPs associated with asthma or liver disease between Estonia, HapMap populations and the 1000 genome

project populations. **International journal of immunogenetics**, v. 46, n. 2, p. 49-58, 2019.

RICKS-SANTI, L. et al. Next generation sequencing reveals high prevalence of BRCA1 and BRCA2 variants of unknown significance in early-onset breast cancer in African American women. **Ethnicity & disease**, v. 27, n. 2, p. 169, 2017.

ROBERTS, M. R. et al. Single-nucleotide polymorphisms in DNA repair genes and association with breast cancer risk in the web study. **Carcinogenesis**, v. 32, n. 8, p. 1223-1230, 2011.

ROMANOWICZ, H. et al. New single nucleotide polymorphisms (SNPs) in homologous recombination repair genes detected by microarray analysis in Polish breast cancer patients. **Clinical and experimental medicine**, v. 17, n. 4, p. 541-546, 2017.

RUAN, Y. et al. Genetic polymorphisms in AURKA and BRCA1 are associated with breast cancer susceptibility in a Chinese Han population. **The Journal of pathology**, v. 225, n. 4, p. 535-543, 2011.

SANTOS, L. S. et al. The role of CCNH Val270Ala (rs2230641) and other nucleotide excision repair polymorphisms in individual susceptibility to well-differentiated thyroid cancer. **Oncology reports**, v. 30, n. 5, p. 2458-2466, 2013.

SCHILDKRAUT, J. M. et al. Association between DNA damage response and repair genes and risk of invasive serous ovarian cancer. **PLoS One**, v. 5, n. 4, 2010.

SECOLIN, R. et al. Distribution of local ancestry and evidence of adaptation in admixed populations. **Scientific reports**, v. 9, n. 1, p. 1-12, 2019.

SHEN, H. et al. Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. **Nature communications**, v. 4, n. 1, p. 1-10, 2013.

SILVERMAN, E.K. Haplotype thinking in lung disease. **Proceedings of the American Thoracic Society**, v. 4, n. 1, p. 4-8, 2007.

SJAARDA, J. et al. Influence of Genetic Ancestry on Human Serum Proteome. **The American Journal of Human Genetics**, 2020.

SLATKIN, M. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. **Nature Reviews Genetics**, v. 9, n. 6, p. 477-485, 2008.

SMETANA, K. et al. Ageing as an important risk factor for cancer. **Anticancer research**, v. 36, n. 10, p. 5009-5017, 2016.

SOLODSKIKH, S. A. et al. Targeted sequencing to discover germline variants in the BRCA1 and BRCA2 genes in a Russian population and their association with breast cancer risk. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**, v. 813, p. 51-57, 2019.

SOLTYS, D.T. et al. Novel XPG (ERCC5) mutations affect DNA repair and cell survival after ultraviolet but not oxidative stress. **Human mutation**, v. 34, n. 3, p. 481-489, 2013.

SONG, H. et al. Common variants in RB1 gene and risk of invasive ovarian cancer. **Cancer research**, v. 66, n. 20, p. 10220-10226, 2006.

SONG, X. et al. Single nucleotide polymorphisms of nucleotide excision repair pathway are significantly associated with outcomes of platinum-based chemotherapy in lung cancer. **Scientific reports**, v. 7, n. 1, p. 1-11, 2017.

SONG, Y. et al. Association between five polymorphisms in vascular endothelial growth factor gene and urinary bladder cancer risk: A systematic review and meta-analysis involving 6671 subjects. **Gene**, 2019.

SPRAGUE, B. L. et al. Genetic variation in TP53 and risk of breast cancer in a population-based case-control study. **Carcinogenesis**, v. 28, n. 8, p. 1680-1686, 2007.

TANG, W. et al. Aurora-A V57I (rs1047972) polymorphism and cancer susceptibility: a meta-analysis involving 27,269 subjects. **PloS one**, v. 9, n. 3, 2014.

TOFOLI, F.A. et al. Increasing the genetic admixture of available lines of human pluripotent stem cells. **Scientific reports**, v. 6, n. 1, p. 1-7, 2016.

WANG, B. et al. The association of six polymorphisms of five genes involved in three steps of nucleotide excision repair pathways with hepatocellular cancer risk. **Oncotarget**, v. 7, n. 15, p. 20357, 2016.

WANG, K. et al. Tracking human population structure through time from whole genome sequences. **PLoS genetics**, v. 16, n. 3, p. e1008552, 2020.

WANG, N. et al. Association of p21 SNPs and risk of cervical cancer among Chinese women. **BMC cancer**, v. 12, n. 1, p. 589, 2012.

WANG, Xinjia; LIU, Zhenyu. Systematic meta-analysis of genetic variants associated with osteosarcoma susceptibility. **Medicine**, v. 97, n. 38, 2018.

WELLER, M. et al. Consanguineous unions and the burden of disability: A population-based study in communities of Northeastern Brazil. **American Journal of Human Biology**, v. 24, n. 6, p. 835-840, 2012.

WEN, J; XU, Q; YUAN, Y. Single nucleotide polymorphisms and sporadic colorectal cancer susceptibility: a field synopsis and meta-analysis. **Cancer cell international**, v. 18, n. 1, p. 155, 2018.

World Health Organization. Global Health Observatory. Geneva: World Health Organization; 2018. who.int/gho/database/en/. Accessed May 21, 2020.

XU, G. P. et al. The association between BRCA1 gene polymorphism and cancer risk: a meta-analysis. **Oncotarget**, v. 9, n. 9, p. 8681, 2018.

XUE, J. et al. Genetic polymorphisms in cyclin D1 are associated with risk of renal cell cancer in the Chinese population. **Oncotarget**, v. 8, n. 46, p. 80889, 2017.

YANG, Fan et al. Association of excision repair cross-complimentary group 1 gene polymorphisms with breast and ovarian cancer susceptibility. **Journal of cellular biochemistry**, v. 120, n. 9, p. 15635-15647, 2019.

YANG, W. et al. Polymorphisms in the 5' upstream regulatory region of p21 WAF1/CIP1 and susceptibility to oesophageal squamous cell carcinoma. **Scientific reports**, v. 6, n. 1, p. 1-9, 2016.

YANG, Y. et al. Association between vascular endothelial growth factor gene polymorphisms and bladder cancer risk. **Molecular and clinical oncology**, v. 2, n. 4, p. 501-505, 2014.

YIN, J. et al. HapMap-based study identifies risk sub-region on chromosome 19q13. 3 in relation to lung cancer among Chinese. **Cancer epidemiology**, v. 37, n. 6, p. 923-929, 2013.

YU, D.D. et al. A review on hepatocyte nuclear factor-1beta and tumor. **Cell & bioscience**, v. 5, n. 1, p. 58, 2015.

YU, T. et al. Rs3212986 polymorphism, a possible biomarker to predict smoking-related lung cancer, alters DNA repair capacity via regulating ERCC1 expression. **Cancer medicine**, v. 7, n. 12, p. 6317-6330, 2018.

YUE, L. L. et al. Association of ATM and BMI-1 genetic variation with breast cancer risk in Han Chinese. **Journal of cellular and molecular medicine**, v. 22, n. 7, p. 3671-3678, 2018.

ZARE, M. et al. Association of MMP-7 -181a>g polymorphism with colorectal cancer and gastric cancer susceptibility: a systematic review and meta-analysis. **ABCD, arq. bras. cir. dig.**, São Paulo, v. 32, n. 3, e1449, 2019.

ZHANG, G. et al. p53 protein expression affected by TP53 polymorphism is associated with the biological behavior and prognosis of low rectal cancer. **Oncology Letters**, v. 18, n. 6, p. 6807-6821, 2019.

ZHANG, Y. et al. Association of single nucleotide polymorphisms in ERCC2 gene and their haplotypes with esophageal squamous cell carcinoma. **Tumor Biology**, v. 35, n. 5, p. 4225-4231, 2014.

ZHANG, Z. et al. Association between the XPG gene rs2094258 polymorphism and risk of gastric cancer. **Journal of clinical laboratory analysis**, v. 32, n. 8, p. e22564, 2018.

ZHAO, J. et al. XPG rs17655 G> C polymorphism associated with cancer risk: evidence from 60 studies. **Ageing (Albany NY)**, v. 10, n. 5, p. 1073, 2018.

ZHOU, Q.Y., et al. Simultaneous multiple single nucleotide polymorphism detection based on click chemistry combined with DNA-encoded probes. **Chemical science**, v. 9, n. 13, p. 3335-3340, 2018.

ZHU, G. et al. Association of nineteen polymorphisms from seven DNA repair genes and the risk for bladder cancer in Gansu province of China. **Oncotarget**, v. 7, n. 21, p. 31372, 2016.

ZHU, J. et al. Association studies of ERCC1 polymorphisms with lung cancer susceptibility: a systematic review and meta-analysis. **PloS one**, v. 9, n. 5, 2014.

ZHU, L. X. et al. 3'-UTR polymorphism (rs10434) in the VEGF gene is associated with B-CLL in a Chinese population. **Genetics and Molecular Research**, v. 14, n. 2, p. 4085-4089, 2015.

ZIYATDINOV, A. et al. Mixed-model admixture mapping identifies smoking-dependent loci of lung function in African Americans. **European Journal of Human Genetics**, p. 1-13, 2019.

APÊNDICES

APÊNDICE A - Revisão de literatura dos 90 SNPs identificados nos 73 idosos longevos de Brejo dos Santos.

Gene	SNP	Alelos	Classif.	Tipo de Câncer	População	Referência
Variantes com alelos iguais a literatura						
<i>BRCA1</i>	rs16942	T/C	Risco	Câncer de mama	Europeus / Asiáticos/ Americanos	(COX et al., 2011)
	rs4986852	C/T	Risco	Câncer de mama	Europeus	(ROMANOWICZ et al., 2017)
	rs2048718	C/T	Protetor	Câncer Cervical	Asiáticos	(DI LIU et al., 2018)
	rs799923	G/A	Risco	Câncer de mama	Asiáticos	(HAN, 2017)
<i>BRCA2</i>	rs4987117	C/T	Risco	Câncer de mama	Europeus/ Asiáticos	SOLODSKIKH et al., 2019).
	rs144848	A/C	Risco Protetor	Linfoma Não-Hodgkin e Câncer de mama Câncer de Pulmão	Caucasianos/ Asiáticos / Africanos Asiáticos	(LI et al., 2017) (LIN et al., 2016)
	rs4987047	A/T	Risco	Câncer de mama	Europeus	(JOHNSON et al., 2007)
<i>TP53</i>	rs1042522	G/C	Risco	Câncer de Cabeça e Pescoço Cervical cancer	Europeus População subsaariana	(FERNÁNDEZ-MATEOS et al, 2019) (KAMIZA et al., 2020)
	rs2078486	G/A	Risco	Câncer de ovário	Não-hispânicos	(SCHILDKRAUT et al., 2010)
<i>AURKA</i>	rs6024836	G/A	Risco	Câncer de mama	Asiáticos	(RUAN et al., 2011)
<i>CCND1</i>	rs3862792	C/T	Risco	Câncer de Próstata Leucemia	Europeus	(RASHKIN et al, 2019)
	rs614367	C/T	Risco	Câncer de Mama	Europeus	(LAMBRECHTS, 2012)
<i>CDKN1A</i>	rs2395655	A/G	Risco	Cacinoma esofágico	Asiáticos	(YANG et al., 2016)

<i>ATM</i>	rs3092856	C/T	Risco	Câncer de mama	Asiáticos	(YUE et al., 2018).
<i>XRCC1</i>	rs1799782	G/A	Risco	Câncer de mama	Asiáticos	(ALIMU et al., 2018
	rs25487	T/C	Protetor	Câncer de mama	Asiáticos	(ZHU et al., 2016)
	rs762507	T/C	Risco	Carcinoma esofágico	Asiáticos	(DAI et al., 2019)
<i>ERCC5</i>	rs17655	G/C	Risco	Câncer gástrico Câncer colorretal	Asiáticos /Caucasianos	(ZHAO et al., 2018)
	rs2227869	G/C	Protetor	Câncer de estômago Câncer de Tireóide	Asiáticos Europeus	(HUSSAIN et al., 2009) (SANTOS et al., 2013)
	rs4150386	A/C	Risco	Câncer endometrial	Americanos	(DOHERTY et al., 2011)
	rs2094258	C/T	Risco	Câncer gástrico	Asiáticos	(ZHANG et al., 2018)
	rs2296148	C/T	Risco	Câncer de Próstata	Afro-americanos	(HOOKER et al., 2008)
	rs4150351	A/C	Protetor	Câncer de Próstata	Americanos	(BARRY et al., 2012)
	rs4150383	G/A	Protetor	Carcinoma Hepatocelular	Asiáticos	(WANG et al., 2016)
	rs4150393	A/G	Risco	Carcinoma espinocelular de cabeça e pescoço	Não hispânicos	(MA et al., 2012)
<i>VEGF</i>	rs1005230	C/T	Pior prognóstico	Glioma e glioblastoma	Asiáticos	(LINHARES et al., 2018)
	rs25648	C/T	Risco	Câncer de Bexiga	Europeus / Asiáticos	(SONG et al., 2019)
	rs3025035	C/T	Risco	Câncer de mama	European	(FURRIOL et al., 2015)
	rs3025039	C/T	Risco	Câncer de mama	Asiáticos	(SONG et al., 2019)

	rs10434	A/G	Risco	Câncer gástrico	Asiáticos	(ZHU et al., 2015)
	rs3025040	C/T	Protetor	Câncer de Tireóide	Asiáticos	(LIU et al., 2017)
	rs833052	C/A	Risco	Câncer de Bexiga	Asiáticos	(YANG et al., 2014)
<i>MMP7</i>	rs10895304	A/G	Risco	Câncer de mama	Asiáticos	(BEEGHLIY-FADIEL et al., 2008)
	rs12285347	C/T	Risco	Câncer de Próstata	Americanos	(HOFFMANN et al., 2017)
<i>ERCC1</i>	rs2336219	G/A	Risco	Câncer colorretal	Asiáticos	(DAI et al., 2015)
	rs6509214	G/T	Risco	Câncer de Pulmão	Asiáticos	(LEE et al., 2015)
	rs10415949	A/G	Risco	Câncer de Pulmão	Asiáticos	(LEE et al., 2015)
	rs3212948	G/C	Risco	Câncer de Pulmão	Asiáticos	(ZHU et al., 2014)
	rs3212986	C/A	Risco	Câncer de pulmão de células não pequenas Câncer de Pâncreas Câncer de ovário Câncer gástrico	Asiáticos	(YU et al., 2018) (HE et al., 2016) (YANG et al., 2019) (HE et al., 2018)
<i>ERCC2</i>	rs238416	T/C	Risco	Adenocarcinoma de Pulmão	Asiáticos	(HAN, 2017)
	rs13181	T/G	Risco	Câncer pulmão Glioma Câncer de laringe	Asiáticos Asiáticos Europeus	(MININA, 2019) (GAO, 2016) (FERNÁNDEZ-MATEOS, 2019)
<i>RBI</i>	rs2854344	G/A	Protetor	Câncer de mama Câncer de Ovário	Asiáticos	(LESUEUR et al., 2006) (SONG et al., 2006)
<i>HNFB1B</i>	rs7405776	G/A	Risco	Câncer de Ovário	Americanos	(SHEN et al., 2013)

	rs1016990	G/C	Risco	Câncer de Próstata	Europeus	(BERNDT, et al., 2011)
	rs11651052	G/A	Risco	Câncer endometrial	Europeus	(PAINTER et al., 2015)
	rs11658063	G/C	Protetor	Câncer de Pulmão	Afro-americanos	(JONES et al., 2019)
	rs11651755	T/C	Risco	Câncer de ovário	Europeus	(BURGHAUS et al 2017)
	rs2005705	G/A	Risco	Câncer endometrial	Caucasianos / Asiáticos	(PAINTER et al., 2015)
	rs7501939	C/T	Protetor	Câncer de Próstata	Europeus	(NIKOLIĆ et al., 2014)
<i>NCOA3</i>	rs2076546	A/G	Protetor	Câncer de mama	Europeus	(BURWINKELet al., 2005)
	rs6094752	C/T	Risco	Câncer de mama	Caucasianos/ Afro-americanos	(HARTMAIER et al., 2009)
<i>CDHI</i>	rs8056538	G/A	Risco	Câncer colorretal	Europeus	(CARVAJAL-CARMONA et al., 2011)
	rs7195409	G/A	Risco	Câncer de pulmão de células não pequenas	Asiáticos	(LI et al., 2018)
	rs7188750	G/A	Protetor	Câncer de mama	Asiáticos	(BEEGHLY-FADIEL et al., 2010)
	rs6499199	C/T	Risco	Câncer endometrial	Asiáticos	(GENG et al., 2018)
	rs4783689	C/T	Protetor	Câncer endometrial	Asiáticos	(GENG et al., 2018)
	rs4782726	A/G	Risco	Câncer de Próstata	Afrodescendentes	(CHANG et al., 2011)
	rs2113200	T/A	Risco	Câncer colorretal	Europeus	(CARVAJAL-CARMONA et al, 2011)
	rs17715799	A/T	Risco	Câncer endometrial	Asiáticos	(GENG et al., 2018)
	rs12919719	C/G	Risco	Câncer de mama	Asiáticos	(BEEGHLY-FADIEL et al., 2010)

Variantes com diferentes alelos						
<i>BRCA1</i>	rs1799950	T/C	Protetor	Câncer de mama	Caucasianos	(XU et al., 2018) Metanálise
	rs4986764	G/A	Protetor	Câncer de mama	Asiáticos	(MA; WU; MA, 2019) Metanálise
	rs4986765	T/C	Risco	Câncer de Mama	Asiáticos	(MA; WU; MA, 2019) Metanálise
<i>BRCA2</i>	rs206119	G/A	Risco	Carcinoma Nasofaríngeo	Asiáticos	(QIN et al., 2011) Caso-controle plataforma de genotipagem GoldenGate (Illumina Inc.), GoldenGate Genotyping (GGGT), (SPRAGUE et al., 2007) caso-controle/n diz
<i>TP53</i>	rs2909430	C/T	Protetor	Câncer de Mama Leucemia	Americanos	(BILOUS et al., 2017) – estudo caso -conjunto de dados do Projeto 1000 Genomes
			Risco		Europeus	
<i>AURKA</i>	rs1047972	T/C	Protetor	Câncer de Mama Câncer Gástrico	Caucasianos	(TANG et al., 2014) - Metanálise (MESIC et al., 2019) – estudo caso-controle
			Risco		Europeus	
<i>CCND1</i>	rs1944129	C/T	Risco	Câncer de rins	Asiáticos	(XUE et al., 2017) – caso-controle
<i>CDKN1A</i>	rs1321311	C/A	Risco	Carcinoma Esofágico	Asiáticos	(GENG et al., 2015)- caso-controle
<i>XRCC1</i>	rs25489	C/T	Risco Risco	Câncer de Mama Câncer de Pâncreas	Europeus / Asiáticos Americanos	(LOIZIDOU et al., 2008) caso-controle (CHEN; WANG; LI, 2019)- METANÁLISE
	rs25486	C/T	Risco Risco	Câncer de Mama Câncer de Pulmão	Americanos Latinos	(ROBERTS et al., 2011) Caso-controle (CHANG et al., 2009) Caso-controle
<i>ERCC5</i>	rs4150360	C/T	Risco	Câncer de Pulmão	Asiáticos	(SONG et al., 2017) Estudo caso
<i>MMP7</i>	rs17098236	C/T	Risco	Câncer de Ovário	Não-hispânicos	(JOHNATT et al., 2010) caso-controle Illumina GoldenGate assay.

	rs11568818	T/C	Risco Risco	Câncer de Próstata Câncer gástrico	Americanos Asiáticos	(BIAŁKOWSKA et al., 2018) caso- controle (ZARE et al., 2019) revisão sistemática
<i>ERCC1</i>	rs1046282	A/G	Risco	Carcinoma Hepatocelular	Asiáticos	(HUANG et al., 2018) caso-controle
	rs3212980	T/G	Risco Risco	Câncer de Pulmão Câncer de Ovário	Asiáticos	(YIN et al., 2013) caso-controle/hapmap (BAO et al., 2020) caso-controle
	rs11615	A/G	Risco	Câncer de Mama	Asiáticos	(LI et al., 2018)
<i>ERCC2</i>	rs1052555	G/A	Pior prognóstico	Câncer colorretal	Asiáticos	(LI et al., 2020)
	rs1799793	C/T	Risco	Câncer Gástrico Carcinoma esofágico	Asiáticos Caucasianos	(LI et al., 2018) (BOLDRIN et al., 2019) coorte/caso
	rs11878644	T/C	Risco	Câncer de Pulmão	Latinos/ Afro- americanos	(CHANG et al., 2008) hapmap
<i>RBI</i>	rs2227311	A/G	Protetor	Câncer de Ovário	Americanos/ Europeus	(SONG et al., 2006)- Illumina BeadStation 500G Golden Gate/hapmap
<i>HNF1B</i>	rs3094509	G/A	Risco	Câncer de Próstata	Asiáticos	(BERNDT, et al., 2011)- painel de ensaio Illumina iSelect personalizado como parte do terceiro estágio do CGEMS
<i>CDH1</i>	rs4076177	A/G	Risco	Câncer de Próstata	Europeus	(GONZÁLEZ et al., 2013) coorte
Variantes com um alelo em comum						
<i>TP53</i>	rs2287497	G/A	Risco	Câncer de Ovário	Europeus	(MEDREK et al., 2013)
	rs12951053	A/C	Risco	Câncer Retal	Asiáticos	(ZHANG 2019)
<i>CDKN1A</i>	rs1801270	C/A	Protetor	Câncer cervical	Asiáticos	(WANG et al., 2012)
<i>ATM</i>	rs4987943	A/G	Protetor	Câncer de Mama	American	(CONCANNON et al., 2008)

<i>XRCC1</i>	rs2307177	T/G	Risco	Câncer de cabeça e pescoço	Asiáticos	(MAHJABEEN et al., 2013)
<i>ERCC2</i>	rs1618536	T/C	Risco	Carcinoma espinocelular de esôfago	Asiáticos	(ZHANG et al., 2014)
<i>CDHI</i>	rs8055236	G/T	Risco	Câncer de Próstata	Asiáticos	(OH et al., 2015)
Alelos não reportados na literatura						
	rs1799949	G/A	Idade do diagnóstico	Câncer de Mama	Afro-americanos	(RICKS-SANTI et al., 2017)

APÊNDICE B - Frequência genotípica dos 90 SNPs e o p-valor de Hardy Weinberg.

Gene	SNP	Genótipo			H. W. p - value
		Wild-type	Heterozigose	Homozigose	
<i>BRCA1</i>	rs16942	TT (34) 46,57%	TC (35) 47,94%	CC (4) 5,47%	0.188
	rs1799950	TT (62) 84,93%	TC (10) 13,69%	CC (1) 1,36%	0.431
	rs4986852	CC (72) 98,63%	CT (1) 1,36%	TT (0) 0%	0.953
	rs1799949	GG (34) 46,57%	AG (35) 47,94%	AA (4) 5,47%	0.188
	rs4986764	GG (22) 30,13%	AG (39) 53,42%	AA (12) 16,43%	0.447
	rs4986765	TT (7) 9,58%	CT (37) 50,68%	CC (29) 39,72%	0.325
<i>BRCA2</i>	rs2048718	CC (20) 27,39%	TC (43) 58,90%	TT (10) 13,69%	0.086
	rs799923	GG (49) 67,12%	AG (21) 28,76%	AA (3) 4,10%	0.695
	rs4987117	CC (71) 97,26%	TC (2) 2,73%	TT (0) 0%	0.905
	rs144848	AA (33) 45,20%	AC (35) 47,94%	CC (5) 6,84%	0.288
	rs4987047	AA (67) 91,78%	AT (4) 5,47%	TT (0) 0%	0.807
	rs206119	GG (9) 12,32%	AG (24) 32,87%	AA (40) 54,79%	0.091
<i>TP53</i>	rs1042522	GG (7) 9,58%	CG (30) 41,09%	CC (34) 46,57%	0.9190
	rs2909430	CC (1) 1,36%	TC (23) 31,50%	TT (49) 67,12%	0.3469
	rs2078486	GG (55) 75,34%	AG (17) 23,28%	AA (1) 1,36%	0.8069
	rs2287497	GG (37) 50,68%	AG (32) 43,83%	AA (3) 4,10%	0.2217
	rs12951053	AA (57) 78,08%	AC (15) 20,54%	CC (1) 1,36%	0.9906
<i>AURKA</i>	rs1047972	TT (3) 4,10%	CT (18) 24,65%	CC (52) 71,23%	0.3813
	rs6024836	GG (30) 41,09%	AG (32) 43,83%	AA (10) 13,69%	0.7549
<i>CCND1</i>	rs3862792	CC (62) 84,93%	TC (10) 13,69%	TT (0) 0%	0.5265
	rs1944129	CC (20) 27,39%	TC (36) 49,31%	TT (17) 23,28%	0.9181
	rs614367	CC (58) 79,45%	TG (13) 17,80%	TT (2) 2,73%	0.2504
<i>CDKN1A</i>	rs1801270	CC (57)	AC (15)	AA (0)	0.3238

		78,08%	20,54%	0%	
	rs2395655	AA (18) 24,65%	AG (42) 57,53%	GG (12) 16,43%	0.1379
	rs1321311	CC (47) 64,38%	AC (22) 30,13%	AA (4) 5,47%	0.5105
<i>ATM</i>	rs4987943	AA (70) 95,89%	AG (2) 2,73%	GG (0) 0%	0.9048
	rs3092856	CC (71) 97,26%	TC (2) 2,73%	TT (0) 0%	0.9055
<i>XRCCI</i>	rs1799782	GG (66) 90,41%	AG (7) 9,58%	AA (0) 0%	0.6669
	rs25489	CC (66) 90,41%	TC (6) 8,21%	TT (1) 1,36%	0.0776
	rs25487	TT (8) 10,95%	TC (28) 38,35%	CC (37) 50,68%	0.4463
	rs2307177	TT (70) 95,89%	TG (3) 4,10%	GG (0) 0%	0.8577
	rs25486	CC (9) 12,32%	TC (28) 38,35%	TT (35) 47,94%	0.3702
	rs762507	TT (11) 15,06%	TC (35) 47,94%	CC (27) 36,98%	0.9503
<i>ERCC5</i>	rs17655	GG (31) 42,46%	CG (33) 45,20%	CC (9) 12,32%	0.9620
	rs2227869	GG (63) 86,30%	CG (9) 12,32%	CC (1) 1,36%	0.3251
	rs4150386	AA (61) 83,56%	AC (11) 15,06%	CC (0) 0%	0.4828
	rs2094258	CC (50) 68,49%	TC (21) 28,76%	TT (2) 2,73%	0.9078
	rs2296148	CC (60) 82,19%	TC (12) 16,43%	TT (1) 1,36%	0.6571
	rs4150351	AA (67) 93,15%	AC (6) 6,84%	CC (0) 0%	0.6669
	rs4150360	CC (31) 42,46%	TC (35) 47,94%	TT (4) 5,47%	0.1436
	rs4150383	GG (45) 61,64%	AG (23) 32,87%	AA (3) 4,10%	0.9288
	rs4150393	AA (68) 93,15%	AG (5) 6,84%	GG (0) 0%	0.7619
<i>VEGF</i>	rs25648	CC (61) 83,56%	TC (10) 13,09%	TT (2) 2,73%	0.0728
	rs3025035	CC (58) 79,45%	TC (14) 19,17%	TT (1) 1,36%	0.8824
	rs3025039	CC (43) 58,90%	TC (28) 38,35%	TT (1) 1,36%	0.1289
	rs10434	AA (9) 12,32%	AG (29) 39,72%	GG (35) 47,94%	0.4416
	rs3025040	CC (44) 60,27%	TC (27) 36,98%	TT (2) 2,73%	0.3662
	rs833052	CC (64) 87,67%	AC (8) 10,95%	AA (1) 1,36%	0.2277

	rs1005230	CC (30) 41,09%	TC (34) 46,57%	TT (9) 12,32%	0.8943
<i>MMP7</i>	rs11568818	TT (21) 28,76%	TC (37) 50,68%	CC (15) 20,54%	0.8603
	rs10895304	AA (36) 49,31%	AG (34) 46,57%	GG (2) 2,73%	0.0674
	rs12285347	CC (17) 27,39%	TC (36) 49,31%	TT (17) 23,28%	0.9181
<i>ERCC1</i>	rs1046282	AA (40) 54,79%	AG (29) 39,72%	GG (3) 4,10%	0.4220
	rs2336219	GG (56) 76,71%	AG (16) 21,91%	AA (1) 1,36%	0.9055
	rs3212980	TT (46) 63,01%	TG (24) 32,87%	GG (3) 4,10%	0.9530
	rs6509214	GG (35) 47,94%	TG (30) 41,09%	TT (8) 10,95%	0.6828
	rs10415949	AA (40) 54,79%	AG (28) 38,35%	GG (5) 6,84%	0.9734
	rs11615	AA (27) 36,98%	AG (32) 43,83%	GG (14) 19,17%	0.4190
	rs3212948	GG (27) 36,98%	CG (33) 45,20%	CC (13) 17,80%	0.6000
	rs3212986	CC (46) 63,01%	AC (24) 32,87%	AA (3) 4,10%	0.9530
<i>ERCC2</i>	rs13181	TT (50) 68,49%	TG (21) 28,76%	GG (2) 2,73%	0.9078
	rs1052555	GG (55) 75,34%	AG (14) 19,17%	AA (2) 2,73%	0.3569
	rs238416	TT (12) 16,43%	TC (36) 49,31%	CC (25) 34,24%	0.8737
	rs1799793	CC (54) 73,97%	TC (15) 20,54%	TT (2) 2,73%	0.4554
	rs1618536	TT (25) 34,24%	TC (30) 41,09%	CC (18) 24,65%	0.1452
	rs11878644	TT (31) 42,46%	TC (27) 36,98%	CC (14) 19,17%	0.0808
<i>RBI</i>	rs2227311	AA (54) 73,97%	AG (19) 26,02%	GG (0) 0%	0.1971
	rs2854344	GG (63) 86,30%	AG (10) 13,69%	AA (0) 0%	0.5298
<i>HNF1B</i>	rs7405776	GG (27) 36,98%	AG (37) 50,68%	AA (9) 12,32%	0.3767
	rs1016990	GG (35) 47,94%	CG (34) 46,57%	CC (4) 5,47%	0.2876
	rs11651052	GG (13) 17,80%	AG (33) 45,20%	AA (27) 36,98%	0.4041
	rs11658063	GG (30) 41,09%	CG (33) 45,20%	CC (10) 13,69%	0.3630
	rs11651755	TT (11) 15,06%	TC (28) 38,35%	CC (34) 46,57%	0.3424
	rs2005705	GG (13) 17,80%	AG (37) 50,68%	AA (23) 31,52%	0.4315

		17,80%	50,68%	31,50%	
	rs3094509	GG (20) 27,39%	AG (34) 46,57%	AA (19) 26,02%	0.4931
	rs7501939	CC (25) 34,24%	TC (35) 47,94%	TT (13) 17,80%	0.4178
<i>NCOA3</i>	rs2076546	AA (59) 80,82%	AG (14) 19,17%	GG (0) 0%	0.4297
	rs6094752	CC (64) 87,67%	TC (9) 12,32%	TT (0) 0%	0.5746
<i>CDHI</i>	rs8055236	GG (43) 58,90%	TG (22) 30,13%	TT (8) 10,95%	0.0633
	rs7195409	GG (5) 6,84%	AG (31) 42,46%	AA (37) 50,68%	0.6609
	rs7188750	GG (52) 71,23%	AG (19) 26,02%	AA (2) 2,73%	0.8680
	rs6499199	CC (62) 84,93%	TC (11) 15,06%	TT (0) 0%	0.4863
	rs4783689	CC (27) 36,98%	TC (35) 47,94%	TT (11) 15,06%	0.9503
	rs4782726	AA (4) 5,47%	AG (28) 38,35%	GG (41) 56,16%	0.7824
	rs4076177	AA (29) 39,72%	AG (34) 46,57%	GG (10) 13,69%	0.9945
	rs2113200	TT (49) 67,12%	AT (21) 26,76%	AA (3) 4,10%	0.4102
	rs17715799	AA (50) 68,49%	AT (20) 27,39%	TT (3) 4,10%	0.5839
	rs12919719	CC (53) 71,23%	CG (17) 23,28%	GG (3) 4,10%	0.3813
	rs8056538	GG (48) 65,75%	GA (21) 28,76%	AA (4) 5,47%	0.4102

Abreviação: H.W. *p* – value: p-valor do Hardy-weinberg

APÊNDICE C - Frequência do alelo menor (MAF) para os 90 SNPs nas populações do 1000 Genomas (Europeu, Africano, Nativo-americano miscigenado) e população de Brejo dos Santos.

SNP	Alelo Menor	Brejo dos Santos	EUR	AFR	NAT-AMR
rs16942	C	0,2945	0,355	0,218	0,37
rs1799950	C	0,0821	0,06	0,003	0,05
rs4986852	T	0,0068	0,029	0,002	0,02
rs1799949	A	0,2945	0,3	0,21	0,35
rs4986764	A	0,4315	0,562	0,6860	0,74
rs4986765	C	0,3493	0,653	0,941	0,83
rs2048718	T	0,4315	0,444	0,227	0,58
rs799923	A	0,1849	0,227	0,01	0,12
rs4987117	T	0,0136	0,025	0,001	0,02
rs144848	C	0,3082	0,295	0,084	0,3
rs4987047	T	0,0281	0	0,033	0
rs206119	A	0,2876	0,784	0,555	0,73
rs1042522	C	0,3098	0,715	0,331	0,68
rs2909430	T	0,1712	0,858	0,714	0,91
rs2078486	A	0,1301	0,087	0,315	0,17
rs2287497	A	0,2638	0,126	0,66	0,21
rs12951053	C	0,1164	0,098	0,11	0,15
rs1047972	C	0,1643	0,818	0,83	0,89
rs6024836	A	0,3611	0,277	0,732	0,46
rs3862792	T	0,0694	0,033	0,152	0,04
rs1944129	T	0,4794	0,496	0,396	0,39
rs614367	T	0,1164	0,1476	0,1251	0,13
rs1801270	A	0,1041	0,067	0,309	0,25
rs2395655	G	0,4583	0,38	0,741	0,39
rs1321311	A	0,2054	0,217	0,411	0,19
rs4987943	G	0,0138	0,001	0,094	0,01
rs3092856	T	0,0068	0,001	0,058	0,01
rs1799782	A	0,0479	0,052	0,072	0,12
rs25489	T	0,0547	0,059	0,026	0,06
rs25487	C	0,3013	0,634	0,89	0,69
rs2307177	G	0,0205	0,011	0,043	0,03
rs25486	T	0,3194	0,636	0,815	0,68
rs762507	C	0,3904	0,593	0,672	0,78
rs17655	C	0,3493	0,25	0,501	0,27
rs2227869	C	0,0753	0,054	0,039	0,02
rs4150386	C	0,0763	0,101	0,007	0,08
rs2094258	T	0,1712	0,182	0,153	0,23
rs2296148	T	0,0958	0,056	0,13	0,03

rs4150351	C	0,0479	0,148	0,006	0,08
rs4150360	T	0,3071	0,527	0,178	0,34
rs4150383	A	0,2083	0,173	0,073	0,11
rs4150393	G	0,0342	0,094	0,024	0,04
rs1005230	T	0,3561	0,505	0,835	0,66
rs25648	T	0,0958	0,173	0,071	0,14
rs3025035	T	0,1095	0,089	0,321	0,06
rs3025039	T	0,2083	0,121	0,084	0,23
rs10434	G	0,3219	0,54	0,675	0,74
rs3025040	T	0,2123	0,123	0,145	0,24
rs833052	A	0,0684	0,868	0,958	0,85
rs17098236	T	0,1164	0,079	0,1	0,06
rs11568818	C	0,4589	0,439	0,452	0,34
rs10895304	G	0,2638	0,182	0,258	0,26
rs12285347	T	0,4794	0,438	0,564	0,35
rs1046282	G	0,243	0,256	0,39	0,37
rs2336219	A	0,1232	0,132	0,298	0,22
rs3212980	G	0,2054	0,249	0,262	0,35
rs6509214	T	0,315	0,344	0,419	0,44
rs10415949	G	0,4041	0,407	0,726	0,32
rs11615	G	0,4109	0,278	0,963	0,61
rs3212948	C	0,4041	0,352	0,949	0,6
rs3212986	A	0,2054	0,25	0,291	0,35
rs13181	G	0,1712	0,364	0,192	0,21
rs1052555	A	0,1267	0,337	0,073	0,15
rs238416	C	0,4109	0,641	0,949	0,55
rs1799793	T	0,1338	0,359	0,073	0,2
rs1618536	C	0,452	0,549	0,936	0,48
rs11878644	C	0,3819	0,418	0,917	0,43
rs2227311	G	0,1319	0,128	0,042	0,17
rs2854344	A	0,0684	0,058	0,002	0,03
rs7405776	A	0,3698	0,358	0,557	0,29
rs1016990	C	0,4794	0,327	0,088	0,22
rs11651052	A	0,178	0,474	0,643	0,39
rs11658063	C	0,4109	0,392	0,405	0,29
rs11651755	C	0,4657	0,483	0,719	0,4
rs2005705	A	0,178	0,434	0,577	0,37
rs3094509	A	0,2739	0,382	0,754	0,33
rs7501939	T	0,3424	0,394	0,557	0,3
rs2076546	G	0,1301	0,084	0,255	0,09
rs6094752	T	0,0616	0,028	0,075	0,03
rs8056538	A	0,1986	0,294	0,277	0,27
rs8055236	T	0,2602	0,182	0,577	0,12
rs7195409	A	0,2808	0,762	0,415	0,64
rs7188750	A	0,1575	0,17	0,39	0,16

rs6499199	T	0,0753	0,054	0,427	0,11
rs4783689	T	0,3904	0,413	0,028	0,33
rs4782726	G	0,2465	0,803	0,651	0,84
rs4076177	G	0,3698	0,363	0,747	0,31
rs2113200	A	0,1986	0,289	0,228	0,27
rs17715799	T	0,178	0,288	0,205	0,22
rs12919719	G	0,1643	0,286	0,168	0,21

APÊNDICE D - Média e desvio-padrão das MAFs nas populações mundiais (Europeu, Africana e Nativo-americano miscigenado) e população de Brejo dos Santos.

	EUR	AFR	NAT-AMR	Brejo dos Santos
Média	0,306273	0,369612	0,312778	0,227456
Desvio-padrão	0,227706	0,305237	0,243103	0,138556

APÊNDICE E – Haplótipos, ancestralidade e referência dos alelos de haplótipos que modificam o risco de câncer.

Crom.	Gene	Haplótipos/Freq.	Ancest.	SNP/Alelos	SNP	Referência
6p.12	<i>VEGF</i>	CC (69,2%)	EUR	rs1005230 - C/T rs25648 - C/T	rs1005230 – Alelo C	Pior prognóstico de glioma (LINHARES et al., 2018)
		TC (19,4%)	EUR/AFR		rs25648 – Alelo T	Risco para o câncer de bexiga (SONG et al., 2019)
		TT (11,3%)	AFR			
		CCG (50,1%)	EUR/AFR	rs3025039 - C/T rs3025040 - C/T rs10434 - A/G	rs3025039 – Alelo T	Risco câncer de mama (SONG et al., 2019)
		CCA (29,1%)	EUR		rs3025040 – Alelo T	Protetor para o câncer de tireóide (LIU et al., 2017)
		TTG (18,1%)	EUR/AFR		rs10434 – Alelo G	Risco para o câncer colorretal (Jeon et al., 2014) Risco para o osteosarcoma (Wang et al., 2018).
		CTG (2,3%)	EUR/AFR		rs10434 - Alelo A	Risco para leucemia linfocítica crônica (ZHU et al., 2015)
11p.13	<i>MMP7</i>	TT (53,8%)	EUR	rs12285347 – C/T rs11568818 - T/C	rs12285347	Nenhum
		CC (43,6%)	EUR		rs11568818	Nenhum
		CT (2,6%)	EUR			
13p.13	<i>ERCC5</i>	AC (68,3%)	EUR	rs4150351 - A/C rs4150360 - C/T	rs4150351 – Alelo C	Protetor para o câncer de próstata (BARRY et al., 2012)
		AT (26,2%)	EUR		rs4150360	Nenhum
		CT (5,5%)	EUR			
16p.13	<i>CDH1</i>	GT (73,0%)	EUR	rs8056538 - G/A rs2113200 - T/A	rs8056538	Risco para o câncer colorretal (CARVAJAL-CARMONA et al., 2011)
		AA (24,8%)	EUR/AFR		rs2113200	Risco para o câncer colorretal (CARVAJAL-CARMONA et al., 2011)
		AT (2,1%)	EUR			
		CA (78,8%)	EUR	rs12919719 - C/G rs17715799 - A/T	rs12919719 – Alelo C	Protetor para o câncer de mama (BEEGHLY-FADIEL et al., 2010)
		GT (18,4%)	EUR/AFR		rs17715799 – Alelo T	Risco para o câncer endometrial (GENG et al., 2018)
		CT (2,4%)	EUR/AFR			

		GC (50,1%)	EUR	rs7188750 - G/A rs4783689 - C/T	rs7188750 – Alelo A	Risco para o câncer de mama (BEEGHLY-FADIEL et al., 2010)	
		GT (27,5%)	EUR/AFR		rs4783689 – Alelo T	Protetor para o câncer endometrial (GENG et al., 2018)	
		AC (22,2%)	EUR/AFR				
17p13	<i>TP53</i>	ATC (58,4%)	EUR/AFR	rs12951053 - A/C rs2909430 – C/T rs1042522 – G/C	rs12951053	Nenhum	
		CTG (15,0%)	EUR/AFR		rs2909430	Nenhum	
		ACG (14,6%)	EUR		rs1042522 – Alelo G	Risco para o câncer cervical (KAMIZA et al., 2020), carcinoma espinocelular de cabeça e pescoço (FERNÁNDEZ-MATEOS et al., 2019)	
		ATG (11,0%)	EUR/AFR		rs1042522 – Alelo C	Risco para o câncer retal (ZHANG et al., 2019) e câncer colorretal (ELSHAZLI et al., 2020)	
	<i>HNFB1B</i>	CGG (49,3%)	EUR	rs7501939 - C/T rs11651052 - G/A rs11658063 - G/C	rs7501939 – Alelo T	Protetor para o câncer de próstata (NIKOLIĆ et al., 2014)	
		CAG (10,8%)	EUR/NAT		rs11651052 - Alelo A	Risco para o câncer endometrial (PAINTER et al., 2015)	
		TAC (31,5%)	EUR/AFR/ NAT		rs11658063 – Alelo C		Protetor para o câncer de pulmão (JONES et al., 2019)
		TAG (6,8%)	EUR/AFR				
	<i>BRCA1</i>	TG (69,2%)	EUR	rs16942 – T/C rs1799949 - G/A	rs16942 – Alelo C	Aumenta o risco para o câncer de mama (COX et al., 2011)	
		CG (2,7%)	EUR/AFR		rs1799949	nenhum	
		CA (27,9%)	AFR				
		GC (69,5%)	EUR/AFR	rs4986764 - G/A rs4986765 - T/C	rs4986764	Nenhum	
		AC (11,7%)	EUR/AFR		rs4986765	Nenhum	
		AT (18,8%)	EUR				
19p13	<i>XRCCI</i>	CTGC (33,5%)	EUR/AFR	rs25487 – T/C rs25486 – C/T rs1799782 - G/A rs762507 – T/C	rs25487 – Alelo T	Protetor para o câncer de bexiga (ZHU et al., 2016)	
		TCGC (28,0%)	EUR		rs25486	Nenhum	
		CTGT (27,6%)	EUR/AFR		rs1799782 – Alelo A	Risco para o câncer de mama (ALIMU et al., 2018)	
		CTAC (8,2%)	EUR/AFR		rs762507 – Alelo T	Risco de carcinoma esofágico (DAI et al., 2019)	

		CCGC (2,1%)	EUR/AFR			
<i>ERCC2</i>		TG (78,2%)	EUR	rs13181 - T/G rs1052555 - G/A	rs13181 – Alelo G	Risco para o câncer de laringe (FERNÁNDEZ-MATEOS et al., 2019), câncer de pulmão (MININA, 2019) e glioma (GAO, 2016).
		GG (7,0%)	EUR			
		GA (14,8%)	EUR		rs1052555	
<i>ERCC1</i>		AGCT (45,3%)	EUR	rs1046282 - A/G rs2336219 - G/A rs3212986 - C/A rs3212980 - T/G	rs1046282	Nenhum
		AACT (21,5%)	EUR		rs2336219 – Alelo A	Risco para o câncer colorretal (DAI et al., 2019)
		GGAG (28,4%)	EUR/AFR		rs3212986 – Alelo C Alelo A	Risco para o câncer de ovário (YANG et al., 2019) Protetor para o câncer gástrico (He et al., 2018).
		GGCT (3,8%)	EUR/AFR		rs3212980	Nenhum

Abreviações: **Crom** – Cromossomo, **Freq** – Frequência, **Ancest** – Ancestralidade. Alelos de risco destacados em cinza claro e alelos protetores destacados em cinza escuro.