Genetics and Molecular Biology, 45, 1, e20210172 (2022)

DOI: https://doi.org/10.1590/1678-4685-GMB-2021-0172

Research Article
Human and Medical Genetics

Haplotypes of single cancer driver genes and their local ancestry in a highly admixed long-lived population of Northeast Brazil

Steffany Larissa Galdino Galisa[1], Priscila Lima Jacob[1], Allysson Allan de Farias[1,3], Renan Barbosa Lemes[3], Leandro Ucela Alves[1,3], Júlia Cristina Leite Nóbrega[1], Mayana Zatz[3], Silvana Santos[1,2] and Mathias Weller[1,2]

[1]*Universidade Estadual da Paraíba (UEPB), Núcleo de Estudos em Genética e Educação, Programa de Pós-Graduação em Saúde Pública, Campina Grande, PB, Brazil.*
[2]*Universidade Estadual da Paraíba (UEPB), Departamento de Biologia, Campina Grande, PB, Brazil.*
[3]*Universidade de São Paulo (USP), Departamento de Genética e Biologia Evolutiva, São Paulo, SP, Brazil.*

Abstract

Admixed populations have not been examined in detail in cancer genetic studies. Here, we inferred the local ancestry of cancer-associated single nucleotide polymorphisms (SNPs) and haplotypes of a highly admixed Brazilian population. SNP array was used to genotype 73 unrelated individuals aged 80–102 years. Local ancestry inference was performed by merging genotyped regions with phase three data from the 1000 Genomes Project Consortium using RFmix. The average ancestry tract length was 9.12–81.71 megabases. Strong linkage disequilibrium was detected in 48 haplotypes containing 35 SNPs in 10 cancer driver genes. All together, 19 risk and eight protective alleles were identified in 23 out of 48 haplotypes. Homozygous individuals were mainly of European ancestry, whereas heterozygotes had at least one Native American and one African ancestry tract. Native-American ancestry for homozygous individuals with risk alleles for *HNF1B, CDH1*, and *BRCA1* was inferred for the first time. Results indicated that analysis of SNP polymorphism in the present admixed population has a high potential to identify new ancestry-associated alleles and haplotypes that modify cancer susceptibility differentially in distinct human populations. Future case-control studies with populations with a complex history of admixture could help elucidate ancestry-associated biological differences in cancer incidence and therapeutic outcomes.

Keywords: Ancestry, admixed population, cancer driver gene, single nucleotide polymorphism (SNP), haplotype.

Received: June 15, 2021; Accepted: November 17, 2021.

Introduction

Cancer is the second leading cause of death in developing countries, and its incidence is expected to increase by 75% by 2030 because of risk-associated lifestyle behaviors and the aging of the world's population (Bray *et al.*, 2012; Cai and Liu, 2019; Torre *et al.*, 2015). In Brazil, as in other developing countries, the oldest old, those aged ≥80 years, is a rapidly growing population (Shetty, 2012; Mathers *et al.*, 2015; Neumann and Albert, 2018). In 2020, the longevous elderly population is estimated to account for 2% of the Brazilian population, that is, 4,441,000 individuals. The oldest-old represent an adequate model of human longevity to study the adverse effects of progressive aging on cancer (Nolen *et al.*, 2017). Longevous individuals present genetic variants associated with cancer susceptibility, and their phenotype manifestation might depend on their ancestry (Aizer *et al.*, 2014; Jin *et al.*, 2016; Özdemir and Dotto, 2017). It is well-established in literature that not only socio-economic but also biological differences can contribute to distinct cancer susceptibilities among human populations. Despite a comparable socio-economic status and lifestyle, Hispanics and Asians in the USA have an overall significantly decreased cancer susceptibility compared to Afro-Americans (Özdemir and Dotto, 2017). Especially the incidence of prostate cancer and triple-negative breast cancer (TNBC) are significantly increased among Afro-American men and women if compared to other populations (Newman *et al.*, 2017, Jiagge *et al.*, 2018; Lewis and Cropp, 2020). The hormone receptor positive subtypes of breast cancer in contrast, are more common among women of European origin compared to Afro-American women (Agboola *et al.*, 2012; Newman *et al.*, 2019)

Germline mutations in tumor suppressor genes, oncogenes, and DNA repair genes have been extensively investigated in genome-wide association studies (GWAS) using European populations (Haiman and Stram, 2010; Park *et al.*, 2018). Cancer prediction based on genomic data frequently uses polygenic-risk scores; however, the predictive ability is lower for populations of different ancestry than those of European descent (Martin *et al.*, 2019). Only 1% of cancer GWAS are performed in African and Latin American populations (Bodian *et al.*, 2014; Fernandes *et al.*, 2016; Park *et al.*, 2018). Genomic variants associated with cancer are often characterized by an ancestry-specific effect called "flip-flop", in which variants associated with cancer in one ancestral population may have no or the opposite association as a result of linkage and epistatic effects (Wang *et al.*, 2018).

Comparative studies of distinct human populations have revealed differences in mutated allele frequency in cancer

Send correspondence to Silvana Santos. Universidade Estadual da Paraíba (UEPB), Núcleo de Estudos em Genética e Educação, Programa de Pós-Graduação em Saúde Pública, Rua Domitila Cabral de Castro, s/n, 3º andar, sala 310, 58.429-570, Campina Grande, PB, Brazil. E-mail: silvanasantos@servidor.uepb.edu.br

driver genes, including oncogenes and tumor suppressor genes (Özdemir and Dotto, 2017; Nakshatri *et al.*, 2019; Bandlamudi and Taylor, 2020; Carrot-Zhang *et al.*, 2020). For example, a South American case-control study identified 13 polymorphisms in a Colombian population that modified the risk of breast cancer, whereas an increase in the proportion of Native Americans decreased the risk of disease (Torres *et al.*, 2019). Local ancestry inference has been used to increase the potential of GWAS through admixture-mapping analysis in ancestrally- diverse populations (Freedman *et al.*, 2006; Yang *et al.*, 2011). Up to date, few studies have inferred local ancestry for cancer causative mutations or identified novel ancestry-associated molecular features (Pasaniuc *et al.*, 2011, Carrot-Zhang *et al.*, 2020; Dutil *et al.*, 2019; Ostrom *et al.*, 2020; Yuan *et al.*, 2018). Ancestry studies of polymorphisms that modify the risk of cancer in admixed populations may help identify genetic differences between populations. It is of special interest that recent studies identified genetic polymorphisms that might have suffered positive selection in native populations and at the same time modify cancer risk (Voskarides, 2018). Data indicated that specific alleles of the *PHD2* gene that are beneficial to hypoxic adaptation also increased the risk of lung cancer among Tibetans (Amorim *et al.*, 2017). The *FADS1* and *FADS2* genes that are involved in fatty acid metabolism and are adaptive for a lipid-rich diet of Siberian Eskimos and Inuit were also suspected to increase the risk of colorectal cancer (Voskarides, 2018). This indicates that ancestry-related genetic polymorphisms can help to elucidate the differences of cancer susceptibility in distinct human populations.

Because the locus-specific ancestry of cancer genomic variants in diverse populations remains unknown, in this study, we inferred the local ancestry of known cancer-associated single nucleotide polymorphisms (SNPs) and haplotypes in a highly admixed population of longevous individuals from Northeast Brazil. This region was chosen because of its high levels of admixture among European settlers, Native Americans, and enslaved Africans (Salzano and Sans, 2014; Moura *et al.*, 2015; Mychaleckyj *et al.*, 2017; de Farias *et al.*, 2018) and increased, as well as a high prevalence of consanguineous marriages compared with other regions of Brazil (Weller *et al.*, 2012).

We analyzed 35 cancer-associated SNPs of 10 genes, and performed a systematic review of the literature to identify the risk and protective alleles of haplotypes. The aim of the study was to identify new haplotypes harboring cancer-associated alleles and their corresponding ancestries. For this purpose, we investigated the frequency of different ancestries in haplotypes with alleles that may modify the risk of cancer and its etiology. A healthy elderly population of individuals aged ≥80 years without a history of cancer was analyzed to determine the proportions of protective and risk alleles in their haplotypes.

Material and Methods

Study population

This cross-sectional study used population genomics methods to analyze 73 unrelated individuals, including 38 women and 35 men aged 80–102 years.

None of the participants was diagnosed with cancer during or before sampling. The elderly samples were obtained from the longitudinal cohort study "Health, Wellbeing and Aging" ["SABE project"] (Lebrão and Laurenti, 2005), which began in 2000 in São Paulo and was extended as a census-based study of elderly individuals aged >60 years from consanguineous communities in the Northeastern Brazil backlands. The "SABE - São Paulo" (SABE-SP) cohort comprises exomic variants of 609 elderly Brazilians available in ABraOM (Online Archive of Brazilian Mutations), a public variant repository (Naslavsky *et al.*, 2017).

The samples were collected in the municipality of Brejo dos Santos in the backlands of the state of Paraíba, Brazil. This community is located at 360 km from Natal and 404 km from João Pessoa, which is at a considerable distance from these capitals of Rio Grande do Norte and Paraíba, respectively, both situated at the Atlantic coast. According to the Brazilian Institute of Geography and Statistics (IBGE), 876 (415 men and 461 women) of the 6198 inhabitants of Brejo dos Santos are >60-years-old. The present cohort represents approximately 10% of the total elderly (≥60-years-old) population of Brejo dos Santos. Of 878 unions in this community, 171 (19.48%) have a consanguineous background, resulting in a coefficient of endogamy of 0.00504 as reported previously (Weller *et al.*, 2012).

The data sampling protocol and consent procedure were reviewed and approved by the National Committee for Ethics in Research (CONEP; Brazil) and are registered under protocol number 0359.0.133.000-11. Written informed consent was obtained from each participant. Consent to publish data anonymously was obtained from each participant.

Selection of genes and corresponding SNPs

Cancer driver genes were identified from studies published between 2005 and 2020 using PubMed. The search terms used were "Oncogenes AND tumor suppressor genes"; and "Genes AND cancer". Eligible studies were those reporting the frequency of cancer risk alleles, haplotypes associated with cancer risk, and ancestry.

The following 20 genes were selected: the tumor suppressor genes *BRCA1*, *BRCA2*, *TP53*, *CDH1*, *ATM*, *MC1R*, *RB1*, and *VEGF* (Song *et al.*, 2006; Al-Moundhri *et al.*, 2010; Lahtz and Pfeifer, 2011; Zhao *et al.*, 2012; Carraro *et al.*, 2013; Tagliabue *et al.*, 2018). The oncogenes *AURKA*, *CCND1*, *NCOA3*, *HNF1B*, *MMP7*, *MITF*, and *CDKN1A* (Burwinkel *et al.*, 2005; Polistena *et al.*, 2014; Hartman and Czyz, 2015; Yu *et al.*, 2015; Vargas-Torres *et al.*, 2016; Tang *et al.*, 2017; Abel *et al.*, 2018); and the DNA repair genes *XRCC1*, *ERCC1*, *ERCC2*, *ERCC5*, and *MLH1* (Xue *et al.*, 2015; Meng *et al.*, 2017).

DNA extraction and genotyping

DNA was extracted from 5 mL of peripheral blood from each individual using the phenol-chloroform method. The quantity and quality of DNA were assessed by gel electrophoresis and spectrophotometry using Nanodrop ND-1000 (Thermo Scientific, Massachusetts, USA). The Axiom® Genome-Wide LAT 1 Array (Affymetrix, USA) was used for genotyping the 73 longevous individuals following the manufacturer's instructions. The array comprises 813,551 SNPs.

Before the analysis, the raw archives of the 73 genotyped individuals were downloaded from Thermo Fisher cloud. SNPs were identified using the software Genotyping Console™ (Version 4.2; Affymetrix Inc). The following filters were used for quality control: dish quality control ≥0.82, quality control call rate ≥92, average call rate for passing ≥97, and minor allele cut off ≥2. A SNP call rate >97% and a Hardy-Weinberg equilibrium of p >0.05 were applied. After filtering, 805,712 SNPs were included.

The minor allele frequency (MAF) of the "SABE Paraíba" (SABE-PB) population was compared with that of the SABE-SP population to determine the allele frequency distribution in two populations of long-lived elderly people living in different geographic regions. MAF information for the SABE-SP population was obtained from the Database of the Brazilian Online Archive of Mutations (ABraOM) (Naslavsky *et al.*, 2017).

The allelic frequency in diverse populations was estimated using the European (N = 1,006), African (N = 1,322), and admixed Native American (N = 694) reference populations from the deep catalog of human genetic variation - *1000 Genomes Project* obtained from the database of Single Nucleotide Polymorphisms using the GRCh37/hg19 reference assembly. The public database of human genetic variants and their relationship with human disease, *Clinvar*, was used to investigate the physical position (hg19), genetic function, classification, and clinical significance of SNPs (Landrum *et al.*, 2016).

Of the 805,712 SNPs, 2,948 were associated with cancer genes covered by the LAT array. Of these, 90 SNPs were selected for the analysis and classified based on the Ensembl genome browser predicted consequence as intronic variant, upstream and downstream variants, untranslated regions (3′-UTR; 5′-UTR), and synonymous or nonsynonymous variants. Cancer susceptibility variants were not identified for *MC1R, MLH1*, and *MITF* in the LAT array, and the remaining 17 genes were therefore included in the allele frequency analysis.

Linkage disequilibrium (LD) of haplotypes and local ancestry inference

Haplotype block identification was performed by calculating the LD of each sequence combination for the 73 genotyped subjects using Haploview 4.0 software (Barrett *et al.*, 2005). Haplotypes were filtered for ≥5% of the MAF. Haplotype analysis was performed by inferring LD blocks. The paired LD structure was built with all SNPs evaluated for each chromosome. Haplotypes that were considerably frequent or rare in the population had frequencies of >20% and <5%, respectively.

For local ancestry inference, the haplotypes of all genetic variants of the elderly and the 1000 genomes project (1KGP) database were analyzed using SHAPEIT2 software (Delaneau *et al.*, 2013), and 1,092 samples from 1KGP Phase 3 were used as a reference panel (The 1000 Genomes Project Consortium *et al.*, 2015). The local ancestry inference (LAI) was estimated using the software RFMix (Maples *et al.*, 2013)**.** We used 0.2 cM and two iterations of expectation maximization with the PopPhased option based on a standard forward-backward algorithm.

The 1KGP database, including Yoruba representing Africans, Iberians representing Europeans, and Peruvians representing admixed Native Americans, was used as described previously (de Farias *et al.*, 2018).

The SNP data were subjected to a cleaning process in which markers with >1% missing genotypes, large deviations from Hardy-Weinberg proportions ($p \leq 10^{-8}$), and MAF <0.01 were excluded, resulting in a final set of 667,855 SNPs after merging. The local ancestry inference was then performed using data from 328 individuals, including 73 from Brejo dos Santos (SABE-PB) and 255 from 1KGP.

Results

Allele frequency

Of 90 selected SNPs (GRCh37 - hg19), 38 were located in intronic regions between exons and two were located in intergenic regions; three were non-coding transcript exon variants; there were eight synonymous and 21 non-synonymous (missense) substitutions; two were nonsense (stop gain), three downstream, six upstream, three in the 5′-UTR, and four in the 3′-UTR (Table S1). Of the 90 genetic variants associated with cancer susceptibility found in the Brejo dos Santos population, 60 had alleles identical to those described in the literature, seven had at least one allele in common, and 23 had different alleles that were not reported previously (Table S2).

The average MAF of the 90 SNPs was 27.7% (range, 0.68–47.94%) (SD = 0.139), whereas the frequency of the same 90 SNPs listed in 1KGP, was 36.96% (SD= 0.305), 30.63% (SD= 0.228), and 31.28% (SD= 0.2431) in populations of African, Caucasian, and Native American ancestry, respectively (Table 1; Table S2). A summary of the protective and risk cancer-associated alleles is shown in Text S1.

Local ancestry inference and ancestral haplotype lengths

In the Brejo dos Santos population, haplotypes containing protective or high-risk alleles were predominantly of European ancestry (Figure S1). For the *HNFB1* gene, 47 (64.38%) haplotypes were of European ancestry, 14 (19.17%) European / African, six (8.21%) European / admixed Native American, two haplotypes (2.73%) African and two (2.73%) African / admixed Native American, one haplotype (1.36%) European / Unknown and one (1.36%) African / Unknown. For the *BRCA1* gene, 37 (50.68%) haplotypes were of European descent, 22 (30.13%) European / African, five (6.84%) European / admixed Native American, five haplotypes (6.84%) were Africans, two (2.73%) African / admixed Native American and one haplotype (1.36%) of admixed Native American ancestry.

The individuals with homologous haplotypes of two different ancestries showed at least one tract of Native American or African origin. The combination of haplotypes with different ancestries was observed with high frequency, and some of them contained homozygous genotypes.

For the tumor suppressor gene *BRCA1*, the frequency of homologous haplotypes with only one ancestry was 28.77–32.88% for European and 1.34–4.11% for African ancestry, and 2.74% were of Native American ancestry (Table 1). Table 1 shows the range of frequencies of homozygous genotypes within homologous haplotypes with the same ancestry.

Table 1 – Frequencies (%) of homozygous genotypes in haplotypes with equal ancestry.

Gene	European	African	Native American
BRCA1	28.77 - 32.88	1.34 - 4.11	2.74
CDH1	31.51 - 41.53	-	2.74
TP53	24.66 - 47.95	4.11 - 5.48	-
VEGF	38.36 - 54.79	2.74 - 6.85	-
HFN1B	32.88 - 38.36	1.37 - 2.74	2.74
MMP7	34.25 - 35.65	-	-
XRCC1	31.51 - 35.62	4.11 - 6.85	-
ERCC1/ERCC2/ERCC5	28.77 - 58.90	2.74 - 6.85	-

The average ancestry tract length was 9.12–81.71 megabases (MB) (Table S3). Longer tracts were observed for populations of European ancestry (30.49–81.71 MB), and variable shorter lengths were observed for African (10.48–26.63 MB) and Native American (9.1–44.94 MB) populations. One of the three ancestries was present in at least one haplotype among ten genes from elderly genotyped data. The average ancestral tract length of *VEGF* for Native American population was almost one half (44.94 MB) of that for the European population (81.71 MB), whereas for *BRCA1*, the African (26.63 MB) stretches were longer. The genes on chromosome 19 showed similar average sizes. Only *BRCA1* had a haplotype breakdown on continuous ancestry considering the four SNPs located close and within the gene (Table S3).

Linkage disequilibrium and alleles of haplotypes that modify cancer risk

In the linkage disequilibrium (LD) analysis, 35 SNPs showed high LD. On chromosome 6 (*VEGF* gene), two blocks of linkage disequilibrium were identified (Figure S2a), on chromosome 11 (*MMP7* gene), one block (Figure S2b), on chromosome 13 (*ERCC5* gene), one block (Figure S2c), on chromosome 16 (*CDH1* gene) three blocks (Figure S2d), on chromosome 17 (*P53*, *HNF1B* and *BRCA1* genes) four blocks (Figure S2e), and on chromosome 19 (*XRCC1*, *ERCC2* and *ERCC1* genes) three LD blocks (Figure S2f). On chromosome 20 no LD blocks were identified (*NCOA3* and *AURKA* genes) (Figure S2g).

Of the 21 haplotypes with strong negative LD ($D < 0$), three were in *BRCA1,* two in *TP53*, and six in *CDH1* (Table 2); four haplotypes corresponded to the *VEGF* tumor suppressor gene and four to the *HNF1B* oncogene. Each of the DNA repair genes *XRCC1* and *ERCC1* had one haplotype with negative LD. Of the haplotypes with strong positive LD ($D > 0$), 11 were in tumor suppressor genes: three in *BRCA1*, two in *TP53*, three in *CDH1*, and three in *VEGF*. Thirteen were in DNA repair genes (Table 2): four in *XRCC1,* three in *ERCC1,* three in *ERCC2,* and three in *ERCC5*. Three haplotypes with strong positive LD were in the *MMP7* oncogene.

A literature search revealed that 25 of the 48 discovered haplotypes did not contain any allele with a known cancer risk modification function (Table 2). The search identified 19 risk alleles in 16 haplotypes and eight protective alleles in eight haplotypes (Table 2). The haplotype ACG of the *TP53* gene contained one risk and one protective allele (Table 2). The haplotype TTG of the *VEGF* gene contained three risk alleles (Table 2). The haplotype CT of the CDH1 gene contained two risk alleles (Table 2). Of the 16 haplotypes with risk alleles, eight (50.0%) had positive LD and eight (50.0%) had negative LD (Table 2). Of the eight haplotypes with protective alleles, two (25.0%) had positive LD and six (75%) had negative LD (Table 2).

Discussion

Data on ancestry-specific variation may improve our understanding of the cancer risk associated with tumor suppressor genes, oncogenes, and DNA repair genes in diverse populations. The present local ancestry results for alleles and haplotypes in elderly individuals identified 48 haplotypes with strong LD that were not previously reported in the literature and alleles that have not been described as risk modifiers. In addition, it identified Native-American and African haplotypes harboring potential cancer variants that were not described previously.

Despite the strong level of endogamy, the present population showed a high degree of admixture: for nine of 10 genes, homozygous haplotypes were present in more than one of the three ancestries. One possible explanation for this result is that many of the haplotypes identified had more than one ancestry. However, most of the genes had combinations of haplotypes of different ancestries, such as single African or European ancestry and combinations of African/European ancestry. This indicated that endogamic effects dominated the present population only after admixture, namely, during the colonization of this region in Northeast Brazil.

To the best of our knowledge, the 48 haplotypes identified in this study have not been described previously in the literature. Their potential function in cancer remains unknown, especially for the 25 haplotypes that do not bear any known allele associated with the modification of cancer risk. All the study participants were ≥80-years-old and did not have a history of cancer, suggesting that the 48 haplotypes included many protective alleles and few risk alleles. However, contrary to this hypothesis, of eight haplotypes containing one protective allele each, six had strong negative LD, and eight of the 16 haplotypes containing at least one risk allele had strong positive LD. Despite these arguments, the present Brazilian elderly population could serve as a model to describe ancestry-specific variation of haplotypes in cancer driver genes of different populations.

Table 2 – The frequency of haplotypes with strong linkage disequilibrium (D), their corresponding SNPs, and genetic locus are shown. The logarithm of the odds ratio of the probability (LOD), the correlation coefficient between loci (r2), and the confidence intervals (CI) are shown for each haplotype. In the SNP column the order of named loci from the top to the bottom corresponds to the order of alleles in the haplotype column from the left to the right hand side. Alleles of haplotypes that increased the risk of cancer in previous studies are highlighted in grey, whereas protective alleles are highlighted in grey and written in bold-type letters.

Genetic locus		Haplotype	SNP	D	LOD	r2	CI	References
6p.12	*VEGF*	CC (69.2%)	**rs1005230**	-0.97	28.18	0.279	0.89-1.0	**rs1005230** (Linhares *et al.*, 2018)
		TC (19.4%)	**rs25648**	-0.97	28.18	0.279	0.89-1.0	**rs25648** (Song *et al.*, 2019)
		TT (11.3%)		-0.97	28.18	0.279	0.89-1.0	
		CCG (50.1%)	**rs3025039**	-0.97	79.85	0.842	0.93-1.0	**rs3025039** (Hou *et al.*, 2017; Wang *et al.*, 2018; Song *et al.*, 2019)
		CCA (29.1%)	**rs3025040**	1	12.6	0.092	0.86-1.0	**rs3025040** (Jeon *et al.*, 2014; Liu *et al.*, 2017)
		TTG (18.1%)	**rs10434**	1	13.93	0.105	0.88-1.0	**rs10434** (Jeon *et al.*, 2014; Zhu *et al.*, 2015; Wang *et al.*, 2018)
		CTG (2,3%)		1	13.93	0.105	0.88-1.0	
11p.13	MMP7	TT (53,8%)	**rs12285347**	1	121.84	0.901	0.98-1.0	**rs12285347** (Hoffmann *et al.*, 2017)
		CC (43.6%)	**rs11568818**	1	121.84	0.901	0.98-1.0	Increased PSA
		CT (2.6%)		1	121.84	0.901	0.98-1.0	**rs11568818** (Beeghly- Fadiel *et al.*, 2009; Sharma *et al.*, 2012; Wu *et al.*, 2013; Wieczorek *et al.*, 2014; Kesh *et al.*, 2015; Horvat *et al.*, 2017; Xie *et al.*, 2016; Bialkowska *et al.*, 2018)
13p.13	*ERCC5*	AC (68,3%)	**rs4150351**	1	13.26	0.126	0.85-1.0	**rs4150351** (Barry *et al.*, 2012; Ma *et al.*, 2012)
		AT (26.2%)	**rs4150360**	1	13.26	0.126	0.85-1.0	**rs4150360** (Song *et al.*, 2017).
		CT (5.5%)		1	13.26	0.126	0.85-1.0	Gastrointestinal toxicity
16p.13	*CDH1*	GT (73,0%)	**rs8056538 rs2113200**	1	101.74	0.895	0.97-1.0	**rs8056538, rs2113200** (Beeghly-Fadiel *et al.*, 2010; Carvajal- Carmona *et al.*, 2011). Colorectal cancer risk
		AA (24.8%)		1	101.74	0.895	0.97-1.0	
		AT (2.1%)		1	101.74	0.895	0.97-1.0	
		CA (78.8%)	**rs12919719**	-0.97	82.37	0.838	0.93-1.0	**rs12919719** (Beeghly-Fadiel *et al.*, 2010); **rs17715799** (Jia *et al.*, 2015; Geng *et al.*, 2018).
		GT (18.4%)	**rs17715799**	-0.97	82.37	0.838	0.93-1.0	
		CT (2.4%)		-0.97	82.37	0.838	0.93-1.0	
		GC (50.1%)	**rs7188750**	-0.96	13.52	0.103	0.82-1.0	**rs7188750** (Beeghly-Fadiel *et al.*, 2010); **rs4783689** (Jia *et al.*, 2015; Geng *et al.*, 2018)
		G**T** (27.5%)	**rs4783689**	-0.96	13.52	0.103	0.82-1.0	
		AC (22.2%)		-0.96	13.52	0.103	0.82-1.0	
17p13	*TP53*	**A**TC (58.4%)	**rs12951053**	1	3.56	0.035	0.58-1.0	**rs12951053** (Mechanic *et al.*, 2007; Ru *et al.*, 2015; Bilous *et al.*, 2016)
		CTG (15.0%)	**rs2909430**	1	28.29	0.258	0.93-1.0	**rs2909430** (Mechanic *et al.*, 2007; Bilous *et al.*, 2017)
		ACG (14.6%)	**rs1042522**	-0.97	25.27	0.258	0.89-1.0	**rs1042522** (Mechanic *et al.*, 2007; Ru *et al.*, 2015; Asai *et al.*, 2019; Fernández- Mateos *et al.*, 2019; Zhang *et al.*, 2019; Ozola *et al.*, 2019; Pouladi *et al.*, 2019; Elshazli *et al.*, 2020; Kamiza *et al.*, 2020; Liu *et al.*, 2020)
		ATG (11.0%)		-0.95	72.4	0.606	0.9-0.98	

Table 2 – Cont.

Genetic locus		Haplotype	SNP	D	LOD	r2	CI	References
17p13	*HNF1B*	CGG (49.3%)	**rs7501939**	-0.95	72.4	0.606	0.9-0.98	**rs7501939** (Setiawan *et al.*, 2012; Chornokur, 2013b; Nikolić *et al.*, 2014; Kristiansen *et al.*, 2015; Ríos-Tamayo *et al.*, 2016; *Oh et al.*, 2017b; Tong *et al.*, 2018)
		CAG (10.8%)	**rs11651052**	-0.95	72.4	0.606	0.9-0.98	
		TA**C** (31.5%)	**rs11658063**	-0.96	79.96	0.688	0.92-0.99	**rs11651052** (Painter *et al.*, 2015).
		TAG (6,8%)		-0.95	51.56	0.453	0.89-0.99	Endometrial cancer risk
								rs11658063 (Jones *et al.*, 2019)
	BRCA1	TG (69.2%)	**rs16942**	-0.99	101.43	0.868	0.96-1.0	**rs16942** (Cox *et al.*, 2011; Heramb *et al.*, 2015; Sagna *et al.*, 2019)
		CG (2.7%)	**rs1799949**	-0.99	101.43	0.868	0.96-1.0	**rs1799949** (Ricks- Santi *et al.*, 2017).
		CA (27.9%)		-0.99	101.43	0.868	0.96-1.0	Age at diagnosis
		GC (69.5%)	**rs4986764**	1	56.42	0.526	0.96-1.0	**rs4986764** (Ma *et al.*, 2013; Ren *et al.*, 2013; Shi *et al.*, 2013; Oussalah *et al.*, 2017; Liu *et al.*, 2018)
		AC (11.7%)	**rs4986765**	1	56.42	0.526	0.96-1.0	**rs4986765** (Oussalah *et al.*, 2017)
		AT (18.8%)		1	56.42	0.526	0.96-1.0	
19p13	*XRCC1*	CTGC (33.5%)	**rs25487**	-0.97	104.68	0.879	0.94-1.0	**rs25487** (Roberts *et al.*, 2011; Jin *et al.*, 2015; Zhu *et al.*, 2016; Alimu *et al.*, 2018; Smolarz and Romanowicz, 2018; Qiao *et al.*, 2018; Minina *et al.*, 2019; Liu *et al.*, 2019; Aboul Enein *et al.*, 2020; Cai *et al.*, 2020; Smolarz *et al.*, 2019)
		TCGC (28.0%)	**rs25486**	1	4.67	0.036	0.67-1.0	
		CTGT (27.6%)	**rs1799782**	1	15.43	0.152	0.89-1.0	**rs25486** (Roberts *et al.*, 2011)
		CTAC (8.2%)	**rs762507**	1	4.99	0.039	0.69-1.0	**rs1799782** (Alimu *et al.*, 2018; Bashir *et al.*, 2018; Li *et al.*, 2018; Zhu *et al.*, 2018; Cai *et al.*, 2020)
		CCGC (2.1%)		1	16.28	0.164	0.9-1.0	**rs762507** (Sacerdote *et al.*, 2013)
								Bladder cancer survival
	ERCC2	TG (78.2%)	**rs13181**	1	59.91	0.622	0.96-1.0	**rs13181** (Dai *et al.*, 2019b; Fernández-Mateos *et al.*, 2019; Li *et al.*, 2019; Smolarz *et al.*, 2019; Balkan *et al.*, 2020; Tavares *et al.*, 2020; Salimzadeh *et al.*, 2020; Zhao *et al.*, 2019)
		GG (7.0%)	**rs1052555**	1	59.91	0.622	0.96-1.0	
		GA (14.8%)		1	59.91	0.622	0.96-1.0	**rs1052555** (Yang *et al.*, 2005; Li *et al.*, 2020)
	ERCC1	AGCT (45.3%)	**rs1046282**	1	9.21	0.132	0.83-1.0	**rs1046282** (Yin *et al.*, 2013)
		AACT (21.5%)	**rs2336219**	-0.95	84.59	0.791	0.91-0.98	**rs2336219** (Ricci *et al.*, 2010; Dai *et al.*, 2019a)
		GGAG (28.4%)	**rs3212986**	1	94.38	0.829	0.97-1.0	**rs3212986** (He *et al.*, 2018; Chaszczewska- Markowska *et al.*, 2019; Gholami *et al.*, 2019; Yang *et al.*, 2019; Bao *et al.*, 2020; Grenda *et al.*, 2020)
		GGCT (3.8%)	**rs3212980**	1	7.82	0.114	0.79-1.0	
								rs3212980 (Yin *et al.*, 2013).

The 72 individuals included in the present study lived under similar socioeconomic conditions within a small community, suggesting that they had similar health-related behaviors such as physical activity, diet, smoking, and consumption of alcohol (Medeiros, 2018). One could speculate that not-sampled individuals of the present population who had cancer, might have other haplotypes with more risk alleles, respectively a lower number of corresponding protective alleles at the analyzed gene loci. However, as we did not perform a case-control study it is impossible with present data to draw any conclusion regarding the potential of haplotypes and corresponding alleles to diminish risk of cancer in the present elderly population.

The high degree of trihybrid ancestry admixture suggests that such a population is a good model for cancer ancestry studies aimed at detecting new haplotypes that represent combinations of polymorphisms with differential effects on the incidence, prognosis, and therapeutic outcome of cancer among human populations. One problem associated with population-specific differences in the etiology, incidence, and prognosis of cancer among individuals can be discriminating between socio-demographic, lifestyle-related factors, and biological differences according to molecular markers (Özdemir and Dotto, 2017). The tumorigenic effect of polymorphisms in cancer driver genes may be enhanced or activated by lifestyle-related risk factors that differ among populations. In this scenario, population-specific molecular differences and lifestyle-related differences are correlated and can lead to meaningful results regarding the molecular differences among populations. On the other hand, lifestyle-related risk factors that differ among populations can lead to the false-positive association of molecular differences that may have no effect on cancer incidence and etiology, or they may mask molecular differences that have distinct biological effects. Many molecular differences among human populations do not affect cancer etiology (Carrot-Zhang *et al.*, 2020). A recent study reported that most molecular differences between African, Asian, and European cancer patients are not limited to tumors, and can be specific to healthy tissues without affecting cancer etiology (Carrot-Zhang *et al.*, 2020).These confounders affecting the identification of molecular cancer-specific differences can be drastically reduced if case-control studies are combined with analysis of ancestry in an admixed population with a relatively homogenous background regarding lifestyle-related risk factors. Particularly with regard to low penetrance polymorphisms, haplotypes could be advantageous over single SNPs in the following aspects: 1. the combination of alleles in a haplotype may have a stronger effect on cancer incidence and etiology; 2. haplotypes with strong LD that contain SNPs with unknown functions may have been under selective pressure and have a specific function; 3. haplotypes with strong LD can also be the result of genetic drift and endogamic effects, and the association of cancer with these evolutionary forces should be analyzed; and 4. haplotype analysis may lead to the identification of SNPs with new functions. In the present study, haplotypes combined SNPs without a known function with SNPs related to cancer risk and etiology.

The complex mosaic derived from the three ancestries revealed a diverse combination of genotypes and ancestries, which reflected the multiple origins of cancer-associated mutations. The *BRCA1* and *CDH1* tumor suppressor genes had a predominant European ancestry and a lower Native American frequency, whereas *TP53* and *VEGF* exhibited a lower African contribution. *BRCA1* is a well-studied gene with a known Native American ancestry based on patient origin and allele frequency studies (Liede *et al.*, 2002; Weitzel *et al.*, 2007), although its presence in haplotypes containing Native American homozygous mutations was described for the first time, which was also the case for *CHD1*. African *BRCA1*, *TP53*, and *VEGF* were found in Brazilian women among other ancestrally diverse populations (Bodian *et al.*, 2014; Fernandes *et al.*, 2016; Oak *et al.*, 2020) and may have a protective effect (Wang *et al.*, 2018).

The *HFN1B* gene, which is associated with prostate cancer, may be ancestry-specific for European Americans and Latinos but not for African-Americans (Waters *et al.*, 2009); the present elderly population presented the Native American haplotypes, which were homozygous. European *MMP7*, which contains cancer-associated mutations, was associated with prostate cancer in European populations in a previous GWAS (Cook *et al.*, 2014). *XRCC1* is associated with breast cancer risk in Mexican admixed individuals, whereas European and African haplotypes did not show cancer-related mutations (Macías-Gómez *et al.*, 2015). The *ERCC* family genes have been associated with esophageal cancer in European patients (Boldrin *et al.*, 2019), and are responsible for the platinum resistance pathway based on cell lines from ancestrally diverse populations (Wheeler *et al.*, 2013).

The flip-flop phenomenon might explain the presence of homozygous individuals for the risk allele with both European and African/Native American haplotypes (Wang *et al.*, 2018). We hypothesized that the flip-flop phenomenon protected longevous individuals. The association between protective alleles and local ancestry inference should be further investigated in Brazilian patients using our dataset as the control in a case control study.

The present study was based on 35 SNPs that can potentially modify the risk of cancer. A literature research revealed that only 10 of these 35 SNPs were described in studies of African or Afro-American populations (Mechanic *et al.*, 2007; Chornokur *et al.*, 2013a; Nikolić *et al.*, 2014; Oh *et al.*, 2017a; Oussalah *et al.*, 2017; Tong *et al.*, 2018; Jones *et al.*, 2019; Sagna *et al.*, 2019; Song *et al.*, 2019; Kamiza *et al.*, 2020), suggesting that most association studies including these SNPs focused on European and Asian populations. The present results showed that a high number of individuals were homozygous for risk alleles with Native American and African ancestry that may modify the risk of cancer. Interestingly, the *HNF1B, CDH1*, and *BRCA1* genes were homozygous for the risk allele that combines Native American, African, or European haplotypes with different frequencies.

Ancestry tracts reconstructed the demographic history of the elderly population of Brejo dos Santos; the haplotypes were broken down to a smaller size over generations as recombination events occurred, and long ancestral tracts were younger than short ones (Pool and Nielsen, 2009; Leitwein *et al.*, 2020). The average ancestry tract length reflected the history of Northeast Brazil, as Native American haplotypes were older than African and European haplotypes. One exception was *VEGF*, which showed a younger African than

Native American tract. This may reflect the origin of the Native American haplotypes after the trans-Atlantic slave trade from West-Africa to the Northeast Brazilian coast. The age of the haplotypes containing cancer-associated mutations remains unknown, and its estimation may help identify Brazilian autochthonous and allochthonous mutations.

An intriguing hypothesis for Native Americans showed a high frequency of cancer mutations associated with low temperatures and high altitude environments where Athabascans and Inuit live (Voskarides, 2018). This "cancer-cold" hypothesis is based on antagonistic pleiotropy effects conferring fitness benefits for SNPs selected under warmer environments such as Brejo dos Santos municipality. More studies are required to assess the extent of this influence with a higher number of municipalities spread by Caatinga semiarid biome.

One important limitation of the present study was the low sample number. Because the number of SNPs for each gene was low, the study may have identified only a small proportion of haplotypes. The potential function of the 48 new haplotypes remains unknown, and we cannot exclude the possibility that endogamic effects and genetic drift generated strong LD of haplotypes without an evolutionary function or contribution to cancer etiology. Another is the use of Peruvians as population reference instead of a Native American from other databases due to incompatibility of coverage and depth with our chosen SNP array.

Conclusions

The present study identified 48 new haplotypes with strong LD and distinct African, Native American, and European ancestry, of which 23 contained alleles that were previously shown to modify the risk of different types of cancer. The results suggested that novel ancestry-specific haplotypes may explain differences in cancer incidence among distinct populations. The present study is the first to identify Native-American ancestry for individuals homozygous for the *HNF1B, CDH1*, and *BRCA1* risk alleles. The results indicated that a high number of haplotypes with the potential to modify cancer risk were associated with African ancestry. African and Native American haplotypes might be associated with increased risk of cancer and also have protective roles. Case-control studies should be performed to elucidate the potential function of the identified haplotypes by comparing the genetic data of healthy controls with those of cancer patients. Studies in an admixed population may help identify haplotypes that contribute to differences in cancer incidence and prognosis in distinct human populations.

Acknowledgements

We would like to thank all men and women of Brejo dos Santos who participated in this study and appreciate the kindly help of the community. This study was funded by following institutions: Universidade Estadual da Paraíba (UEPB- PROPESQ); Fundação de Apoio à Pesquisa do Estado da Paraíba (FAPESQ/CNPq-PPSUS 015/2014); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-INCT14/50931-3); Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-CEPID 2013/08028-1).

Conflict of Interest

The authors declare no potential conflicts of interest.

Author Contributions

SLGG, PLJ, AAF, MW: conception and design, development of methodology, analysis and interpretation of data; LUA, JCLN, SS: acquisition of data and constructing of database; SLGG, PLJ, AAF, MW, RBL, SS, MZ: writing, review, and revision of the manuscript; SS and MW: study supervision. All authors read and approved the final version.

References

1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:7422-7456.

Abel EV, Goto M, Magnuson B, Abraham S, Ramanathan N, Hotaling E, Alaniz AA, Kumar-Sinha C, Dziubinski ML, Urs S *et al.* (2018) HNF1A is a novel oncogene that regulates human pancreatic cancer stem cell properties. eLife 7:e33947.

Aboul Enein AA, Khaled IAA, Khorshied MM, Abdel-Aziz AO, Zahran N, El Saeed AM, Shousha HI and Abdel Rahman HA (2020) Genetic variations in DNA-repair genes (XRCC1, 3, and 7) and the susceptibility to hepatocellular carcinoma in a cohort of Egyptians. J Med Virol 92:3609-3616.

Agboola AJ, Musa AA, Wanangwa N, Abdel-Fatah T, Nolan CC, Ayoade BA, Oyebadejo TY, Banjo AA, Deji-Agboola AM, Rakha EA *et al.* (2012) Molecular characteristics and prognostic features of breast cancer in Nigerian compared with UK women. Breast Cancer Res Treat 135:555-569.

Aizer AA, Wilhite TJ, Chen MH, Graham PL, Choueiri TK, Hoffman KE, Martin NE, Trinh QD, Hu JC and Nguyen PL (2014) Lack of reduction in racial disparities in cancer-specific mortality over a 20-year period. Cancer 120:1532-1539.

Al-Moundhri MS, Al-Khanbashi M, Al-Kindi M, Al-Nabhani M, Burney IA, Al-Farsi A and Al-Bahrani B (2010) Association of E-cadherin (CDH1) gene polymorphisms and gastric cancer risk. World J Gastroenterol 16:3432-3436.

Alimu N, Qukuerhan A, Wang S, Abdurehim Y, Kuyaxi P, Zhang B and Yasheng Y (2018) The association between XRCC1 polymorphism and laryngeal cancer susceptibility in different ethnic groups in Xinjiang, China. Int J Clin Exp Pathol 11:4595-4604.

Amorim CEG, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM and Hünemeier T (2017) Genetic signature of natural selection in first Americans. Proc Natl Acad Sci U S A 114:2195-2199.

Asai T, Tsuchiya Y, Mishra K, Behari A, Shukla P, Ikoma T, Kapoor VK and Nakamura K (2019) Carcinogen Metabolism Pathway and Tumor Suppressor Gene Polymorphisms and Gallbladder Cancer Risk in North Indians: A Hospital-Based Case-Control Study. Asian Pac J Cancer Prev 20: 3643-3647.

Balkan E, Bilici M, Gundogdu B, Aksungur N, Kara A, Yasar E, Dogan H and Ozturk G (2020) ERCC2 Lys751Gln rs13181 and XRCC2 Arg188His rs3218536 Gene Polymorphisms Contribute to Susceptibility of Colon, Gastric, Liver, Lung and Prostate Cancer. J BUON 25:574-581.

Bandlamudi C and Taylor BS (2020) Bridging the Gap: The impact of genetic ancestry on routes to tumorigenesis. Cancer Cell 37:619-621.

Bao Y, Yang B, Zhao J, Shen S and Gao J (2020) Role of common ERCC1 polymorphisms in cisplatin-resistant epithelial ovarian cancer patients: A study in Chinese cohort. Int J Immunogenet 47:443-453.

Barrett JC, Fry B, Maller J and Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265.

Barry KH, Koutros S, Andreotti G, Sandler DP, Burdette LA, Yeager M, Freeman LEB, Lubin JH, Ma X, Zheng T *et al*. (2012) Genetic variation in nucleotide excision repair pathway genes, pesticide exposure and prostate cancer risk. Carcinogenesis 33:331-337.

Bashir K, Sarwar R, Fatima S, Saeed S, Mahjabeen I and Akhtar Kayani M (2018) Haplotype analysis of XRCC1 gene polymorphisms and the risk of thyroid carcinoma. J BUON 23:234-243.

Beeghly-Fadiel A, Shu XO, Long J, Li C, Cai Q, Cai H, Gao YT and Zheng W (2009) Genetic polymorphisms in the MMP-7 gene and breast cancer survival. Int J Cancer 124:208-214.

Beeghly-Fadiel A, Lu W, Gao YT, Long J, Deming SL, Cai Q, Zheng Y, Shu XO and Zheng W (2010) E-cadherin polymorphisms and breast cancer susceptibility: A report from the Shanghai Breast Cancer Study. Breast Cancer Res Treat 121:445-452.

Białkowska K, Marciniak W, Muszyńska M, Baszuk P, Gupta S, Jaworska-Bieniek K, Sukiennicki G, Durda K, Gromowski T, Prajzendanc K *et al*. (2018) Association of zinc level and polymorphism in MMP-7 gene with prostate cancer in Polish population. PLoS One 13:e0201065.

Bilous NI, Abramenko IV, Chumak AA, Dyagil IS and Martina ZM (2016) The distribution of TP53 gene polymorphisms in chronic lymphocytic leukemia patients, sufferers of Chornobyl nuclear power plant accident. Exp Oncol 38:252–256.

Bilous NI, Abramenko I, Saenko V, Chumak A, Dyagil I, Martina Z and Kryachok I (2017) Clinical relevance of TP53 polymorphic genetic variations in chronic lymphocytic leukemia. Leuk Res 58:1-8.

Bodian DL, McCutcheon JN, Kothiyal P, Huddleston KC, Iyer RK, Vockley JG and Niederhuber JE (2014) Germline variation in cancer-susceptibility genes in a healthy, ancestrally diverse cohort: Implications for individual genome sequencing. PLoS One 9:e94554.

Boldrin E, Malacrida S, Rumiato E, Battaglia G, Ruol A, Amadori A and Saggioro D (2019) Association between ERCC1 rs3212986 and ERCC2/XPD rs1799793 and OS in patients with advanced esophageal cancer. Front Oncol 9:85.

Bray F, Jemal A, Grey N, Ferlay J and Forman D (2012) Global cancer transitions according to the Human Development Index (2008-2030): A population-based study. Lancet Oncol 13:790-801.

Burwinkel B, Wirtenberger M, Klaes R, Schmutzler RK, Grzybowska E, Försti A, Frank B, Bermejo JL, Bugert P, Wappenschmidt B *et al.* (2005) Association of NCOA3 polymorphisms with breast cancer risk. Clin Cancer Res 11:2169-2174.

Cai Z and Liu Q (2019) Understanding the Global Cancer Statistics 2018: Implications for cancer control. Sci China Life Sci 64:1017-1020.

Cai W, Liu X, Li Y, Bi B, Liu L and Wang Z (2020) New sights on the associations between the XRCC1 gene polymorphisms and hepatocellular carcinoma susceptibility. J Cell Biochem 121:1005-1022.

Carraro DM, Koike Folgueira MMA, Lisboa BCG, Olivieri EHR, Krepischi ACV, Carvalho AF, Mota LDC, Puga RD, Maciel MS, Michelli RAD *et al.* (2013) Comprehensive analysis of BRCA1, BRCA2 and TP53 germline mutation and tumor characterization: A portrait of early-onset breast cancer in Brazil. PLoS One. 8:e57581.

Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, Zhou W, Berger AC, Huang KL, Newberg JY *et al.* (2020) Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. Cancer Cell 37:639-654.e6.

Carvajal-Carmona LG, Cazier JB, Jones AM, Howarth K, Broderick P, Pittman A, Dobbins S, Tenesa A, Farrington S, Prendergast J *et al*. (2011) 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. Hum Mol Genet 20:2879-2888.

Chaszczewska-Markowska M, Kosacka M, Chryplewicz A, Dyła T, Brzecka A and Bogunia-Kubik K (2019) ECCR1 and NFKB2 Polymorphisms as Potential Biomarkers of Non-small Cell Lung Cancer in a Polish Population. Anticancer Res 39:3269-3272.

Chornokur G, Han G, Tanner R, Lin HL, Lee Green B, Pow-Sang J and Phelan CM (2013a) High grade prostate intraepithelial neoplasia (PIN) is a PSA-independent risk factor for prostate cancer in African American men: Results from a pilot study. Cancer Lett 331:154-157.

Chornokur G, Amankwah EK, Davis SN, Phelan CM, Park JY, Pow-Sang J and Kumar NB (2013b) Variation in HNF1B and obesity may influence prostate cancer risk in african american men: A pilot study. Prostate Cancer 2013:384594.

Cook MB, Wang Z, Yeboah ED, Tettey Y, Biritwum RB, Adjei AA, Tay E, Truelove A, Niwa S, Chung CC *et al.* (2014) A genome-wide association study of prostate cancer in West African men. Hum Genet 133:509-521.

Cox DG, Simard J, Sinnett D, Hamdi Y, Soucy P, Ouimet M, Barjhoux L, Verny-Pierre C, McGuffog L, Healey S *et al*. (2011) Common variants of the BRCA1 wild-type allele modify the risk of breast cancer in BRCA1 mutation carriers. Hum Mol Genet 20:4732-4747.

Dai L, Tao H, Xiong G, Guan X, Bai Y and Xu X (2019a) Association between intronic polymorphisms of XRCC1, ERCC2 and LIG1 genes and risk of esophageal squamous cell carcinoma in a Chinese Han population. Int J Clin Exp Med 12:2710-2719.

Dai P, Li J, Li W, Qin X, Wu X, Di W and Zhang Y (2019b) Genetic polymorphisms and pancreatic cancer risk: A PRISMA-compliant systematic review and meta-analysis. Medicine (Baltimore) 98:e16541.

de Farias AA, Nunes K, Lemes RB, Moura R, Fernandes GR, Melo US, Zatz M, Kok F and Santos S (2018) Origin and age of the causative mutations in KLC2, IMPA1, MED25 and WNT7A unravelled through Brazilian admixed populations. Sci Rep 8:16552.

Delaneau O, Howie B, Cox AJ, Zagury JF and Marchini J (2013) Haplotype estimation using sequencing reads. Am J Hum Genet 93:687-696.

Dutil J, Teer JK, Golubeva V, Yoder S, Tong WL, Arroyo N, Karam R, Echenique M, Matta JL and Monteiro AN (2019) Germline variants in cancer genes in high-risk non-BRCA patients from Puerto Rico. Sci Rep 9:17769.

Elshazli RM, Toraih EA, Elgaml A, Kandil E and Fawzy MS (2020) Genetic polymorphisms of TP53 (rs1042522) and MDM2 (rs2279744) and colorectal cancer risk: An updated meta-analysis based on 59 case-control studies. Gene 734:144391.

Fernandes GC, Michelli RAD, Galvão HCR. Paula AE, Pereira R, Andrade CE, Felicio PS, Souza CP, Mendes DRP, Volc S *et al*. (2016) Prevalence of BRCA1/BRCA2 mutations in a Brazilian population sample at-risk for hereditary breast cancer and characterization of its genetic ancestry. Oncotarget 7:80465-80481.

Fernández-Mateos J, Seijas-Tamayo R, Adansa Klain JC, Pastor Borgoñón M, Pérez-Ruiz E, Mesía R, del Barco E, Salvador Coloma C, Rueda Dominguez A, Caballero Daroqui J *et al*.

(2019). Genetic susceptibility in head and neck squamous cell carcinoma in a Spanish population. Cancers (Basel) 11:493.

Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM *et al.* (2006) Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A 103:14068-14073.

Geng YH, Wang ZF, Jia YM, Zheng LY, Chen L, Liu DG, Li XH, Tian XX and Fang WG (2018) Genetic polymorphisms in CDH1 are associated with endometrial carcinoma susceptibility among Chinese Han women. Oncol Lett 16:6868-6878.

Gholami M, Larijani B, Sharifi F, Hasani-Ranjbar S, Taslimi R, Bastami M, Atlasi R and Amoli MM (2019) MicroRNA-binding site polymorphisms and risk of colorectal cancer: A systematic review and meta-analysis. Cancer Med 8:7477-7499.

Grenda A, Błach J, Szczyrek M, Krawczyk P, Nicoś M, Kuźnar Kamińska B, Jakimiec M, Balicka G, Chmielewska I, Batura-Gabryel H *et al*. (2020) Promoter polymorphisms of TOP2A and ERCC1 genes as predictive factors for chemotherapy in non-small cell lung cancer patients. Cancer Med 9:605-614.

Haiman CA and Stram DO (2010) Exploring genetic susceptibility to cancer in diverse populations. Curr Opin Genet Dev 20:330-335.

Hartman ML and Czyz M (2015) MITF in melanoma: Mechanisms behind its expression and activity. Cell Mol Life Sci 72:1249-1260.

He J, Zhuo ZJ, Zhang A, Zhu J, Hua RX, Xue WQ, Zhang HD, Zhang JB, Li XZ and Jia WH (2018) Genetic variants in the nucleotide excision repair pathway genes and gastric cancer susceptibility in a southern Chinese population. Cancer Manag Res 10:765-774.

Heramb C, Ekstrøm PO, Tharmaratnam K, Hovig E, Møller P and Mæhle L (2015) Ten modifiers of BRCA1 penetrance validated in a Norwegian series. Hered Cancer Clin Pract 13:14.

Hoffmann TJ, Passarelli MN, Graff RE, Emami NC, Sakoda LC, Jorgenson E, Habel LA, Shan J, Ranatunga DK, Quesenberry CP *et al*. (2017) Genome-wide association study of prostate-specific antigen levels identifies novel loci independent of prostate cancer. Nat Commun 8:14248.

Horvat M, Potocnik U, Repnik K, Kavalar R, Zadnik V, Potrc S and Stabuc B (2017) Single nucleotide polymorphisms in genes MACC1, RAD18, MMP7 and SDF-1a as prognostic factors in resectable colorectal cancer. Radiol Oncol 51:151-159.

Hou Q, Li MY, Huang WT, Wei FF, Peng JP, Lou MW and Qiu JG (2017) Association between three VEGF polymorphisms and renal cell carcinoma susceptibility: a meta-analysis. Oncotarget 8:50061-50070.

Jeon YJ, Kim JW, Park HM, Jang HG, Kim JO, Oh J, Chong SY, Kwon SW, Kim EJ, Oh D *et al*. (2014) Interplay between 3′-UTR polymorphisms in the vascular endothelial growth factor (VEGF) gene and metabolic syndrome in determining the risk of colorectal cancer in Koreans. BMC Cancer 14:881.

Jia YM, Xie YT, Wang YJ, Han JY, Tian XX and Fang WG (2015) Association of Genetic Polymorphisms in CDH1 and CTNNB1 with Breast Cancer Susceptibility and Patients' Prognosis Among Chinese Han Women. PLoS One 10:e0135865.

Jiagge E, Chitale D and Newman LA (2018) Triple-negative breast cancer, stem cells, and African ancestry. Am J Pathol 188:271-279.

Jin EH, Kim J, Lee S and Hong JH (2015) Association between polymorphisms in APE1 and XRCC1 and the risk of gastric cancer in Korean population. Int J Clin Exp Med 8:11484-11489.

Jin H, Pinheiro PS, Xu J and Amei A (2016) Cancer incidence among Asian American populations in the United States, 2009-2011. Int J Cancer 138:2136-2145.

Jones CC, Bradford Y, Amos CI, Blot WJ, Chanock SJ, Harris CC, Schwartz AG, Spitz MR, Wiencke JK, Wrensch MR *et al*. (2019) Cross-cancer pleiotropic associations with lung cancer risk in African Americans. Cancer Epidemiol Biomarkers Prev 28:715-723.

Kamiza AB, Kamiza S, Singini MG and Mathew CG (2020) Association of TP53 rs1042522 with cervical cancer in the sub-Saharan African population: A meta-analysis. Trop Med Int Health 25:666-672.

Kesh K, Subramanian L, Ghosh N, Gupta V, Gupta A, Bhattacharya S, Mahapatra NR and Swarnakar S (2015) Association of MMP7− 181A→ G promoter polymorphism with gastric cancer risk: Influence of nicotine in differential allele-specific transcription via increased phosphorylation of cAMP-response element-binding protein (Creb). J Biol Chem 290:14391-14406.

Kristiansen W, Karlsson R, Rounge TB, Whitington T, Andreassen BK, Magnusson PK, Fosså SD, Adami HO, Turnbull C, Haugen TB *et al*. (2015) Two new loci and gene sets related to sex determination and cancer progression are associated with susceptibility to testicular germ cell tumor. Hum Mol Genet 24:4138-4146.

Lahtz C and Pfeifer GP (2011) Epigenetic changes of DNA repair genes in cancer. J Mol Cell Biol 3:51-58.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J *et al.* (2016) ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Res 44:D862-8.

Lebrão ML and Laurenti R (2005) Saúde, bem-estar e envelhecimento: O estudo SABE no Município de São Paulo. Rev Bras Epidemiol 8:127-141.

Leitwein M, Duranton M, Rougemont Q, Gagnaire PA and Bernatchez L (2020) Using haplotype information for conservation genomics. Trends Ecol Evol 35:245-258.

Lewis DD and Cropp CD (2020) The impact of African ancestry on prostate cancer disparities in the era of precision medicine. Genes (Basel) 11:1471.

Li Q, Ma R and Zhang M (2018) XRCC1 rs1799782 (C194T) polymorphism correlated with tumor metastasis and molecular subtypes in breast cancer. Onco Targets Ther 11:8435-8444.

Li W, Zhang M, Huang C, Meng J, Yin X and Sun G (2019) Genetic variants of DNA repair pathway genes on lung cancer risk. Pathol Res Pract 215:152548.

Li YK, Xu Q, Sun LP, Gong YH, Jing JJ, Xing CZ and Yuan Y (2020) Nucleotide excision repair pathway gene polymorphisms are associated with risk and prognosis of colorectal cancer. World J Gastroenterol 26:307-323.

Liede A, Jack E, Hegele RA and Narod SA (2002) A BRCA1 mutation in Native North American families. Hum Mutat 19:460.

Linhares P, Viana-Pereira M, Ferreira M, Amorim J, Nabiço R, Pinto F, Costa S, Vaz R and Reis RM (2018) Genetic variants of vascular endothelial growth factor predict risk and survival of gliomas. Tumour Biol 40:1010428318766273.

Liu R, Ning L, Liu X, Zhang H, Yu Y, Zhang S, Rao W, Shi J, Sun H, and Yu Q (2017) 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 8:15838-15845.

Liu D, Zheng Y, Wang M, Deng Y, Lin S, Zhou L,Yang P, Dai C, Xu P, Hao Q, Song D *et al*. (2018) 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. Aging (Albany NY), 10:266-277.

Liu GC, Zhou YF, Su XC and Zhang J (2019) Interaction between TP53 and XRCC1 increases susceptibility to cervical cancer development: A case control study. BMC Cancer 19:24.

Liu P, Zhuo ZJ, Zhu J, Yang Z, Xin Y, Li S, Li L, Li Y, Wang H and He J (2020) Association of TP53 rs1042522 C> G and miR-34b/c rs4938723 T> C polymorphisms with hepatoblastoma susceptibility: A seven-center case–control study. J Gene Med 22:e3182.

Ma H, Yu H, Liu Z, Wang LE, Sturgis EM and Wei Q (2012) Polymorphisms of XPG/ERCC5 and risk of squamous cell carcinoma of the head and neck. Pharmacogenet Genomics 22:50-57.

Ma XD, Cai GQ, Zou W, Huang YH, Zhang JR, Wang DT and Chen BL (2013) First evidence for the contribution of the genetic variations of BRCA1-interacting protein 1 (BRIP1) to the genetic susceptibility of cervical cancer. Gene 524:208-213.

Macías-Gómez NM, Peralta-Leal V, Meza-Espinoza JP, Gutiérrez-Angulo M, Durán-González J, Ramírez-González JM, Toro AGD, Noberto-Rodriguez A and Leal-Ugarte E (2015) Polymorphisms of the XRCC1 gene and breast cancer risk in the Mexican population. Fam Cancer 14:349-354.

Maples BK, Gravel S, Kenny EE and Bustamante CD (2013) RFMIX: A discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet 93:278-288.

Martin AR, Kanai M, Kamatani Y, Okada Y and Neale BM (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51:584-591.

Mathers CD, Stevens GA, Boerma T, White RA and Tobias MI (2015) Causes of international increases in older age life expectancy. Lancet 385:540-548.

Mechanic LE, Bowman ED, Welsh JA, Khan MA, Hagiwara N, Enewold L, Shields PG, Burdette L, Chanock S and Harris CC (2007) Common genetic variation in TP53 is associated with lung cancer risk and prognosis in African Americans and somatic mutations in lung tumors. Cancer Epidemiol Biomarkers Prev 16:214-222.

Medeiros JB (2018) Fatores associados à hipertensão arterial em idosos longevos residentes em um município do nordeste brasileiro. M. Sc. Thesis, Universidade Estadual da Paraíba, Campina Grande, 50 p.

Meng Q, Wang S, Tang W, Wu S, Gao N, Zhang C, Cao X, Li X, Zhang Z, Aschner M *et al.* (2017) XRCC1 mediated the development of cervival cancer through a novel Sp1/Krox-20 swich. Oncotarget 8:86217-86226.

Minina VI, Bakanova ML, Soboleva OA, Ryzhkova AV, Titov RA, Savchenko YA, Sinitsky MY, Voronina EN, Titov VA and Glushkov AN (2019) Polymorphisms in DNA repair genes in lung cancer patients living in a coal-mining region. Eur J Cancer Prev 28:522-528.

Moura RR de, Coelho AVC, Balbino VQ, Crovella S and Brandão LAC (2015) Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. Am J Hum Biol 27:674-680.

Mychaleckyj JC, Havt A, Nayak U, Pinkerton R, Farber E, Concannon P, Lima AA and Guerrant RL (2017) Genome-wide analysis in Brazilians reveals highly differentiated native American genome regions. Mol Biol Evol 34:559-574.

Nakshatri H, Kumar B, Burney HN, Cox ML, Jacobsen M, Sandusky GE, D'Souza-Schorey C and Storniolo AMV (2019) Genetic ancestry–dependent differences in breast cancer–induced field defects in the tumor-adjacent normal breast. Clin Cancer Res 25:2848-2859.

Naslavsky MS, Yamamoto GL, de Almeida TF, Ezquina SAM, Sunaga DY, Pho N, Bozoklian D, Sandberg TOM, Brito LA, Lazar M *et al.* (2017) Exomic variants of an elderly cohort of Brazilians in the ABraOM database. Hum Mutat 38:751-763.

Neumann LTV and Albert SM (2018) Aging in Brazil. Gerontologist. 58:611-617.

Newman LA and Kaljee LM (2017) Health disparities and triple-negative breast cancer in African American women. JAMA Surg 152:485-493.

Newman LA, Jenkins B, Chen Y, Oppong JK, Adjei E, Jibril AS, Hoda S, Cheng E, Chitale D, Bensenhaver JM *et al.* (2019) Hereditary susceptibility for triple negative breast cancer associated with Western Sub-Saharan African ancestry: Results from an international surgical breast cancer collaborative. Ann Surg 270:484-492.

Nikolić ZZ, Branković AS, Savić-Pavićević DLJ, Preković SM, Vukotić VD, Cerović SJ, Filipović NN, Tomović SM, Romac SP, Brajušković GN *et al.* (2014) Assessment of association between common variants at 17q12 and prostate cancer risk-evidence from Serbian population and meta-analysis. Clin Transl Sci 7:307-313.

Nolen SC, Evans MA, Fischer A, Corrada MM, Kawas CH and Bota DA (2017) Cancer-Incidence, prevalence and mortality in the oldest-old. A comprehensive review. Mech Ageing Dev 164:113-126.

Oak N, Cherniack AD, Mashl RJ, Analysis Network TCGA, Hirsch FR, Ding L, Beroukhim R, Gümüş ZH, Plon SE and Huang KL (2020) Ancestry-specific predisposing germline variants in cancer. Genome Med 12:51.

Oh JJ, Lee SJ, Hwang JY, Kim D, Lee SE, Hong SK, Ho JN, Yoon S, Sung J, Kim WJ *et al.* (2017a) Exome-based genome-wide association study and risk assessment using genetic risk score to prostate cancer in the Korean population. Oncotarget 8:43934-43943.

Oh JJ, Kim TJ, Lee IJ, Song BD, Lee DH, Jung YS, Lee HM, Hong SK, Lee S, Ho JN *et al.* (2017b) MP33-19 An Exome-Wide Association Study Replicated for Prostate Cancer in Korean population. J Urology 197:e424-e425.

Ostrom QT, Egan KM, Nabors LB, Gerke T, Thompson RC, Olson JJ, LaRocca R, Chowdhary S, Eckel-Passow JE, Armstrong G *et al.* (2020) Glioma risk associated with extent of estimated European genetic ancestry in African Americans and Hispanics. Int J Cancer 146:739-748.

Oussalah A, Avogbe PH, Guyot E, Chery C, Guéant-Rodriguez RM, Ganne-Carrié N, Cobat A, Moradpour D, Nalpas B, Negro F *et al*. (2017) BRIP1 coding variants are associated with a high risk of hepatocellular carcinoma occurrence in patients with HCV-or HBV-related liver disease. Oncotarget 8:62842-62857.

Özdemir BC and Dotto GP (2017) Racial differences in cancer susceptibility and survival: More than the color of the skin? Trends Cancer 3:181-197.

Ozola A, Ruklisa D and Pjanova D (2019) The complementary effect of rs1042522 in TP53 and rs1805007 in MC1R is associated with an elevated risk of cutaneous melanoma in Latvian population. Oncol Lett 18:5225-5234.

Painter JN, O'Mara TA, Batra J, Cheng T, Lose FA, Dennis J, Michailidou K, Tyrer JP, Ahmed S, Ferguson K *et al*. (2015) Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. Hum Mol Genet 24:1478-1492.

Park SL, Cheng I and Haiman CA (2018) Genome-wide association studies of cancer in diverse populations. Cancer Epidemiol Biomarkers Prev 27:405-417.

Pasaniuc B, Zaitlen N, Lettre G, Chen GK, Tandon A, Kao WHA, Ruczinski I, Fornage M, Siscovick DS, Zhu X *et al.* (2011) Enhanced statistical tests for GWAS in admixed populations: Assessment using African Americans from CARe and a Breast Cancer Consortium. PLoS Genet 7:e1001371.

Polistena A, Cucina A, Dinicola S, Stene C, Cavallaro G, Ciardi A, Orlando G, Arena R, D'ermo G, Cavallaro A *et al.* (2014) MMP7 expression in colorectal tumours of different stages. In Vivo 28:105-110.

Pool JE and Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics 181:711-719.

Pouladi N, Abdolahi S, Farajzadeh D and Hosseinpour Feizi MA (2019) Haplotype and linkage disequilibrium of TP53-WRAP53 locus in Iranian-Azeri women with breast cancer. PLoS One 14:e0220727.

Qiao L, Feng X, Wang G, Zhou B, Yang Y and Li M (2018) Polymorphisms in BER genes and risk of breast cancer: Evidences from 69 studies with 33760 cases and 33252 controls. Oncotarget 9:16220-16233.

Ren LP, Xian YS, Diao DM, Chen Y, Guo Q and Dang CX (2013) Further evidence for the contribution of the BRCA1-interacting protein-terminal helicase 1 (BRIP1) gene in breast cancer susceptibility. Genet Mol Res 12:5793-5801.

Ricci-Vitiani L, Pallini R, Biffoni M, Todaro M, Invernici G, Cenci T, Maira G, Parati EA, Stassi G, Larocca LM *et al*. (2010) Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. Nature 468:824-828.

Ricks-Santi L, McDonald JT, Gold B, Dean M, Thompson N, Abbas M, Wilson B, Kanaan Y, Naab TJ and Dunston G (2017) Next Generation Sequencing Reveals High Prevalence of BRCA1 and BRCA2 Variants of Unknown Significance in Early-Onset Breast Cancer in African American Women. Ethn Dis 27:169-178.

Ríos-Tamayo R, Lupiañez CB, Campa D, Hielscher T, Weinhold N, Martínez-López J, Jerez A, Landi S, Jamroziak K, Dumontet C *et al*. (2016) A common variant within the HNF1B gene is associated with overall survival of multiple myeloma patients: Results from the IMMEnSE consortium and meta-analysis. Oncotarget 7:59029-59048.

Roberts MR, Shields PG, Ambrosone CB, Nie J, Marian C, Krishnan SS, Goerlitz DS, Modali R, Seddon M, Lehman T *et al*. (2011) Single-nucleotide polymorphisms in DNA repair genes and association with breast cancer risk in the web study. Carcinogenesis 32:1223-1230.

Ru JY, Cong Y, Kang WB, Yu L, Guo T and Zhao JN (2015) Polymorphisms in TP53 are associated with risk and survival of osteosarcoma in a Chinese population. Int J Clin Exp Pathol 8:3198-3203.

Sacerdote C, Guarrera S, Ricceri F, Pardini B, Polidoro S, Allione A, Critelli R, Russo A, Andrew AS, Ye Y *et al*. (2013) Polymorphisms in the XRCC1 gene modify survival of bladder cancer patients treated with chemotherapy. Int J Cancer 133:2004–2009

Sagna T, Bonora E, Ouedraogo MNL, Fusco D, Zoure AA,Cyrille B, Florencia D, Gilbert KJ, Nayi Z, Zoenabo D *et al.* (2019) Identification of BRCA1/2 p.Ser1613Gly, p.Pro871Leu, p.Lys1183Arg, p.Glu1038Gly, p.Ser1140Gly, p.Ala2466Val, p.His2440Arg variants in women under 45 years old with breast nodules suspected of having breast cancer in Burkina Faso. Biomol Concepts 10:120-127.

Salimzadeh H, Lindskog EB, Gustavsson B, Wettergren Y and Ljungman D (2020) Association of DNA repair gene variants with colorectal cancer: risk, toxicity, and survival. BMC Cancer 20:409.

Salzano FM and Sans M (2014) Interethnic admixture and the evolution of Latin American populations. Genet Mol Biol 37:151-170.

Setiawan VW, Haessler J, Schumacher F, Cote ML, Deelman E, Fesinmeyer MD, Henderson BE, Jackson RD, Vöckler JS, Wilkens LR *et al*. (2012) HNF1B and endometrial cancer risk: Results from the PAGE study. PLoS One 7:e30390.

Sharma KL, Misra S, Kumar A and Mittal B (2012) Higher risk of matrix metalloproteinase (MMP-2, 7, 9) and tissue inhibitor of metalloproteinase (TIMP-2) genetic variants to gallbladder cancer. Liver Int 32:1278-1286.

Shetty P (2012) Grey matter: Ageing in developing countries. Lancet 379:1285-1287.

Shi J, Tong J, Cai S, Qu X and Liu Y (2013) Correlation of the BACH1 Pro919Ser polymorphism with breast cancer risk: A literature-based meta-analysis and meta-regression analysis. Exp Ther Med 6:435-444.

Smolarz B and Romanowicz H (2018) Association between single nucleotide polymorphism of DNA repair genes and endometrial cancer: A case-control study. Int J Clinical Exp Pathol 11:1732-1738.

Smolarz B, Michalska MM, Samulak D, Romanowicz H and Wójcik L (2019) Polymorphism of DNA repair genes in breast cancer. Oncotarget 10:527-535.

Song H, Ramus SJ, Shadforth D, Quaye L, Kjaer SK, DiCioccio RA, Dunning AM, Hogdall E, Hogdall C, Whittemore AS *et al.* (2006) Common variants in RB1 gene and risk of invasive ovarian cancer. Cancer Res 66:10220-10226.

Song X, Wang S, Hong X, Li X, Zhao X, Huai C, Chen H, Gao Z, Qian J, Wang J *et al*. (2017) Single nucleotide polymorphisms of nucleotide excision repair pathway are significantly associated with outcomes of platinum-based chemotherapy in lung cancer. Sci Rep 7:11785.

Song Y, Yang Y, Liu L and Liu X (2019) 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 698:186-197.

Tagliabue E, Gandini S, Bellocco R, Maisonneuve P, Newton-Bishop J, Polsky D, Lazovich D, Kanetsky PA, Ghiorzo P, Gruis NA *et al.* (2018) MC1R variants as melanoma risk factors independent of at-risk phenotypic characteristics: A pooled analysis from the M-SKIP Project. Cancer Manag Res 10:1143-1154.

Tang A, Gao K, Chu L, Zhang R and Yang J (2017) Aurora kinases: Novel therapy targets in cancers. Oncotarget 8:23937-23954.

Tavares CB, Alves-Ribeiro FA, Nery Junior EDJ, de Vasconcelos-Valença RJ, Campos-Verdes LC, Gomes FDCSA, Lopes-Costa PV, Santos AR, Pinho-Sobral AL, Campelo V *et al*. (2020) Association of XRCC1 rs1799782 and ERCC2 rs13181 Polymorphisms with Glioma Risk: A Systematic Review and Meta-Analysis. Research Square. DOI: 10.21203/rs.3.rs-56338/v1.

The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526:68-74.

Tong Y, Qu Y, Li S, Zhao F, Wang Y and Mu D (2018) Cumulative evidence for relationships between multiple variants of HNF1B and the risk of prostate and endometrial cancers. BMC Med Genet 19:128.

Torre LA, Bray F, Siegel FL, Ferlay J and Lortet-Tieulent J (2015) Global cancer statistics, 2012. CA Cancer J Clin 65:87-108.

Torres D, Bermejo JL, Mesa KG, Gilbert M, Briceño I, Pohl-Zeidler S, Silos RG, Boekstegers R, Plass C and Hamann U (2019) Interaction between genetic ancestry and common breast cancer susceptibility variants in Colombian women. Int J Cancer 144:2181-2191.

Vargas-Torres SL, Portari EA, Silva AL, Klumb EM, da Rocha Guillobel HC, Camargo MJ, Santos-Rebouças CB, Russomano FB and Macedo JMB (2016) Roles of CDKN1A gene polymorphisms (rs1801270 and rs1059234) in the development of cervical neoplasia. Tumour Biol 37:10469-10478.

Voskarides K (2018) Combination of 247 Genome-Wide Association Studies reveals high cancer risk as a result of evolutionary adaptation. Mol Biol Evol 35:473-485.

Wang S, Qian F, Zheng Y, Ogundiran T, Ojengbede O, Zheng W, Blot W, Nathanson KL, Hennis A, Nemesure B *et al.* (2018) Genetic variants demonstrating flip-flop phenomenon and breast cancer risk prediction among women of African ancestry. Breast Cancer Res Treat 168:703-712.

Water KM, Marchand LL, Kolonel LN, Monroe KR, Stram DO, Henderson BE and Haiman CA (2009) Generalizability of associations from prostate cancer genome-wide association studies in multiple populations. Cancer Epidemiol Biomarkers Prev 18:1285-1289.

Weitzel JN, Lagos VI, Herzog JS, Judkins T, Hendrickson B, Ho JS, Ricker CN, Lowstuter KJ, Blazer KR, Tomlinson G *et al.* (2007) Evidence for common ancestral origin of a recurring BRCA1 genomic rearrangement identified in high-risk Hispanic families. Cancer Epidemiol Biomarkers Prev 16:1615-1620.

Weller M, Tanieri M, Pereira JC, Almeida EDS, Kok F and Santos S (2012) Consanguineous unions and the burden of disability: A population-based study in communities of Northeastern Brazil. Am J Hum Biol 24:835-840.

Wheeler HE, Gamazon ER, Stark AL, O'Donnell PH, Gorsic LK, Huang RS, Cox NJ and Dolan ME (2013) Genome-wide meta-analysis identifies variants associated with platinating agent susceptibility across populations. Pharmacogenomics J 13:35-43.

Wieczorek E, Wasowicz W, Gromadzinska J and Reszka E (2014) Functional polymorphisms in the matrix metalloproteinase genes and their association with bladder cancer risk and recurrence: A mini-review. Int J Urol 21:744-752.

Wu H, Qiao N, Wang Y, Jiang M, Wang S, Wang C and Hu L (2013) Association between the telomerase reverse transcriptase (TERT) rs2736098 polymorphism and cancer risk: Evidence from a case-control study of non-small-cell lung cancer and a meta-analysis. PLoS One 8:e76372.

Xie B, Zhang Z, Wang H, Chen Z, Wang Y, Liang H, Yang G, Yang X and Zhang H (2016) Genetic polymorphisms in MMP 2, 3, 7, and 9 genes and the susceptibility and clinical outcome of cervical cancer in a Chinese Han population. Tumour Biol 37:4883-4888.

Xue M-H, Li GY, Wu XJ, Zhang CX, Zhang CF and Zhu KX (2015) Genetic variability of genes in NER pathway influences the treatment outcome of gastric cancer. Int J Clin Exp Pathol 8:5563-5569.

Yang P, Kollmeyer TM, Buckner K, Bamlet W, Ballman KV and Jenkins RB (2005) Polymorphisms in GLTSCR1 and ERCC2 are associated with the development of oligodendrogliomas. Cancer 103:2363-2372.

Yang JJ, Cheng C, Devidas M, Cao X, Fan Y, Campana D, Yang W, Neale G, Cox NJ, Scheet P *et al.* (2011) Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. Nat Genet 43:237-241.

Yang F, Mu X, Bian C, Zhang H, Yi T, Zhao X and Lin X (2019) Association of excision repair cross-complimentary group 1 gene polymorphisms with breast and ovarian cancer susceptibility. J Cell Biochem 120:15635-15647.

Yin J, Vogel U, Wang H, Ma Y, Wang C, Liang D, Liu J, Yue L, Zhao Y and Ma J (2013) HapMap-based study identifies risk sub-region on chromosome 19q13. 3 in relation to lung cancer among Chinese. Cancer Epidemiol 37:923-929.

Yu D-D, Guo SW, Jing YY, Dong YL and Wei LX (2015) A review on hepatocyte nuclear factor-1beta and tumor. Cell Biosci 5:58.

Yuan J, Hu Z, Mahal BA, Zhao SD, Kensler KH, Jingjiang Pi, Xiaowen Hu, Zhang Y, Wang Y, Jiang J *et al.* (2018) Integrated analysis of genetic ancestry and genomic alterations across cancers. Cancer Cell 34:549-560.e9.

Zhang G, Xu Q, Wang Z, Sun L, Lv Z, Liu J, Xing C and Yuan Y (2019) p53 protein expression affected by TP53 polymorphism is associated with the biological behavior and prognosis of low rectal cancer. Oncol Lett 18:6807-6821.

Zhao Z, Ba C, Wang W, Wang X, Xue R and Wu X (2012) Vascular endothelial growth factor (VEGF) gene polymorphisms and colorectal cancer: A meta-analysis of epidemiologic studies. Genet Test Mol Biomarkers 16:1390-1394.

Zhao Y, Zhao E, Zhang J, Chen Y, Ma J and Li H (2019) A comprehensive evaluation of the association between polymorphisms in XRCC1, ERCC2, and XRCC3 and Prognosis in Hepatocellular Carcinoma: A Meta-Analysis. J Oncol 2019:2408946.

Zhu LX, Ye XJ, Wang YG, Zhu JJ, Xie WZ, Zhao YM and Lai XY (2015) 3'-UTR polymorphism (rs10434) in the VEGF gene is associated with B-CLL in a Chinese population. Genet Mol Res 14:4085-4089.

Zhu G, Su H, Lu L, Guo H, Chen Z, Sun Z, Song R, Wang X, Li H and Wang Z (2016) Association of nineteen polymorphisms from seven DNA repair genes and the risk for bladder cancer in Gansu province of China. Oncotarget 7:31372-31383.

Zhu J, Qi P and Li Z (2018) Interaction Between XRCC1 Gene Polymorphisms and Obesity on Susceptibility to Papillary Thyroid Cancer in Chinese Han Population. Cell Physiol Biochem 49:638-644.

Internet Resources

Health, Wellbeing and Aging (SABE) project, http://hygeia3.fsp.usp.br/sabe/ (accessed 7 May 2020).

Online Archive of Brazilian Mutations (ABraOM), http://abraom.ib.usp.br/index.php (accessed 23 June 2020).

National Committee for Ethics in Research (CONEP), http://conselho.saude.gov.br/web_comissoes/conep/index.html (accessed 3 March 2020).

Clinically relevant variation clinvar (Clinvar), https://www.ncbi.nlm.nih.gov/clinvar/ (accessed 19 June 2020).

Haploview, https://www.broadinstitute.org/haploview/haploview (accessed 21 July 2020).

Supplementary Material

The following online material is available for this article:

Figure S1 – The local ancestry of chromosome 17 using RFMix software.

Figure S2 – Pairwise linkage disequilibrium (LD) graph for chromosomes 6, 11, 13, 16, 17, and 19 generated by Haploview.

Table S1 – Frequencies of the minor and major alleles and corresponding location, genetic function (Ensembl), and clinical significance (Clinvar) of the 90 SNPs for the 73 genotyped samples.

Table S2 – Literature review of the 90 SNPs identified in the Brejo dos Santos population showing the SNPs with the same alleles, different alleles, or at least one common allele found in the literature.

Table S3 – The average ancestry tract lengths in megabases and the respective standard deviations.

Text S1 – Risk alleles and protective alleles of the 48 haplotypes found in the literature.

Associate Editor: Mara H. Hutz