


RESEARCH ARTICLE

Open Access



Haplotype analysis of the internationally distributed *BRCA1* c.3331_3334delCAAG founder mutation reveals a common ancestral origin in Iberia

Anna Marie De Asis Tuazon^{1†}, Paul Lott^{1†}, Mabel Bohórquez^{2†}, Jennyfer Benavides^{2†}, Carolina Ramirez², Angel Criollo², Ana Estrada-Florez², Gilbert Mateus², Alejandro Velez^{3,4}, Jenny Carmona⁴, Justo Olaya⁵, Elisha Garcia¹, Guadalupe Polanco-Echeverry¹, Jacob Stultz¹, Carolina Alvarez⁶, Teresa Tapia⁶, Patricia Ashton-Prolla^{7,8,9}, Brazilian Familial Cancer Network^{7,8,9}, Ana Vega¹⁰, Conxi Lazaro^{11,12}, Eva Tornero^{11,12}, Cristina Martinez-Bouzas¹³, Mar Infante¹⁴, Miguel De La Hoya¹⁵, Orland Diez¹⁶, Brian L. Browning¹⁷, COLUMBUS Consortium, Bruce Rannala¹⁸, Manuel R. Teixeira^{19†}, Pilar Carvallo^{6†}, Magdalena Echeverry^{2†} and Luis G. Carvajal-Carmona^{1,20,21,22*†} 

Abstract

Background: The *BRCA1* c.3331_3334delCAAG founder mutation has been reported in hereditary breast and ovarian cancer families from multiple Hispanic groups. We aimed to evaluate *BRCA1* c.3331_3334delCAAG haplotype diversity in cases of European, African, and Latin American ancestry.

Methods: BC mutation carrier cases from Colombia ($n = 32$), Spain ($n = 13$), Portugal ($n = 2$), Chile ($n = 10$), Africa ($n = 1$), and Brazil ($n = 2$) were genotyped with the genome-wide single nucleotide polymorphism (SNP) arrays to evaluate haplotype diversity around *BRCA1* c.3331_3334delCAAG. Additional Portuguese ($n = 13$) and Brazilian ($n = 18$) BC mutation carriers were genotyped for 15 informative SNPs surrounding *BRCA1*. Data were phased using SHAPEIT2, and identical by descent regions were determined using BEAGLE and GERMLINE. DMLE+ was used to date the mutation in Colombia and Iberia.

Results: The haplotype reconstruction revealed a shared 264.4-kb region among carriers from all six countries. The estimated mutation age was ~ 100 generations in Iberia and that it was introduced to South America early during the European colonization period.

(Continued on next page)

* Correspondence: lgcarvajal@ucdavis.edu

†Anna Marie De Asis Tuazon, Paul Lott, Mabel Bohórquez, Jennyfer Benavides are co-first authors and contributed equally to this work.

†Manuel R. Teixeira, Pilar Carvallo, Magdalena Echeverry and Luis G. Carvajal-Carmona are co-senior authors.

¹Genome Center, University of California Davis, Davis, CA, USA

²⁰Division de Investigaciones, Fundacion de Genética y Genómica, Ibaguè, Colombia

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: Our results suggest that this mutation originated in Iberia and later introduced to Colombia and South America at the time of Spanish colonization during the early 1500s. We also found that the Colombian mutation carriers had higher European ancestry, at the *BRCA1* gene harboring chromosome 17, than controls, which further supported the European origin of the mutation. Understanding founder mutations in diverse populations has implications in implementing cost-effective, ancestry-informed screening.

Keywords: Breast cancer, Haplotype, *BRCA1* c.3331_3334delCAAG, Founder mutation

Background

Breast cancer (BC) remains the most common form of cancer and the second leading cause of cancer death among women and about 5–10% have hereditary breast cancer, explained by genetic susceptibility [1, 2]. Germline mutations in the tumor suppressor gene *BRCA1* account for the largest proportion of BC susceptibility to date and confer a 55–65% lifetime risk of developing breast cancer [2, 3]. *BRCA1* has a very heterogeneous mutation spectrum, often having high frequency of founder mutations in isolated populations such as the Ashkenazi Jewish or the Icelandic population, where few founder mutations account for most *BRCA1* carriers [4, 5].

Among Hispanic populations from Iberia and the Americas, *BRCA1* c.3331_3334delCAAG (Breast Cancer Information Core designation: 3450del4 or rs80357903) is one of the most widely distributed founder mutation and reaches its highest frequency in admixed populations from Central Colombia [6]. *BRCA1* c.3331_3334delCAAG was first described in a Canadian BC family [7], and since then reported in Europe, Latin American, the Middle Eastern, and North African patients [8–15]. The occurrence of *BRCA1* c.3331_3334delCAAG in different populations may be indicative of a mutational hotspot associated with multiple origins or a founder effect from a single ancient mutation. Although haplotype analysis has been carried out for *BRCA1* c.3331_3334delCAAG in some of these countries, they have been limited to a few intragenic markers and to a limited number of populations, often using a single individual from a carrier family [8–10]. Moreover, the *BRCA1* c.3331_3334delCAAG mutation haplotype has not been assessed on an international scale, and the ancestral origin of *BRCA1* c.3331_3334delCAAG remains to be determined. To gain insights into its origin, extensive haplotype analysis of *BRCA1* c.3331_3334delCAAG was completed in carriers from six different countries, and the age of the mutation was estimated in Colombia and Iberia. We utilized genome-wide and targeted SNP data followed by imputation, haplotype phasing, linkage disequilibrium analyses, genetic admixture estimation, and mutation dating to comprehensively assess genetic variation, spanning the entire chromosome 17, where *BRCA1* resides. Our results indicated that *BRCA1* c.3331_3334delCAAG had a single origin in Iberia.

Materials and methods

Study populations

Mutation carriers

The study was carried out using de-identified samples from of 89 *BRCA1* c.3331_3334delCAAG mutation carrier BC cases from Colombia ($n = 32$ cases from Ibaguè and Neiva), Spain ($n = 13$), Portugal ($n = 16$), one of which that originated in Angola (a former Portuguese colony), Chile ($n = 10$), and Brazil ($n = 18$). Mutation carriers were previously ascertained as part of population studies (Colombia, Chile and Brazil) or through high-risk hereditary cancer clinics (Spain and Portugal) [10, 11, 13–16] where all individuals signed informed consent forms and were recruited with locally approved research and clinical testing protocols.

Genotyping and quality control procedures

Array genotyping

Sixty mutation carriers were genotyped with Affymetrix Axiom Human UK Biobank single nucleotide polymorphism (SNP) arrays. Samples with genotyping call rates < 95% were excluded. Basic quality control (for genotypes and missingness per individual) was completed by filtering markers with a genotype rate less than 95%, minor allele frequency ≤ 0.05 , and Hardy-Weinberg equilibrium ≤ 0.00001 . In total, 52 of the 60 samples passed all QC procedures.

Individual SNP genotyping

As additional 31 mutation Brazilian and Portuguese *BRCA1* c.3331_3334delCAAG carriers became available for our study after we completed the SNP genotyping, we decided to carry out targeted genotyping of 15 SNPs around *BRCA1* (seven and eight markers on each side of the gene, Supplementary Table 1) that were informative as they had high heterozygosity, were roughly equally spaced around the minimally shared haplotype, and had high call rates in the SNP arrays. These markers were individually genotyped with the KASP allele-specific genotyping system (LGC Genomics, London, England) following the manufacturer's protocol and in reactions that included non-template controls, two *BRCA1* c.3331_3334delCAAG carriers (positive controls) and two *BRCA1* c.3331_

3334delCAAG non-carriers (mutation negative controls). A summary of mutation carriers and genotype data are detailed in Supplementary Table 2.

Control SNP array data

Data available with the same SNP array on 886 Colombian control matched with cases by sex and geographical origin, were also available for analysis in this study. In addition, for genetic admixture analyses, we used publicly available genotype data from the 1000Genomes study.

Haplotype reconstruction and IBD analysis

All analyses were carried out using GRCh37/hg19 chromosomal positions. Single nucleotide markers (SNPs) on chromosome 17, used to obtain the haplotype that flanks the *BRCA1* c.3331_3334delCAAG mutation, were phased using SHAPEIT [17] with the dataset of 938 (886 controls and 52 mutation carriers that passed genotyping QC) unrelated samples. Following phasing, BEAGLE 4.0 was used for detection of segments that were IBD [18, 19]. The *ibdtrim* parameter, which specifies the number of markers in a 0.15-cM region, was set to 29 for chromosome 17. The lengths of the shared haplotype segments were calculated based on a previous study by Marroni et al. [20], calculated as the sum of the distance to the last marker on either side of the *BRCA1* mutation where all mutation carriers had identical alleles. These IBD segments were verified in parallel using GERMLINE [21] as an alternative approach.

Phylogenetic analysis of mutation haplotypes

The distance from one individual to another was determined by subtracting the distance shared from the length of chromosome 17. A phylogenetic tree was then constructed utilizing the genetic distance between mutation carriers with the UPGMA algorithm, which was incorporated in Clustal Omega [22]. This tool utilizes bootstrap analysis of 1000 replications to assess the statistical confidence in the branching order of the phylogenetic tree. SplitsTree 4.0 was used for visualization (www.splitstree.org/).

Estimating the age of *BRCA1* c.3331_3334delCAAG in Iberia and Colombia

Sixty SNPs in a 4.34-Mb region flanking *BRCA1* (chr17:39040105-43387103) were selected for mutation dating. These markers captured the margins of the different mutation haplotypes determined from IBD analysis, where recombination events were observed. The DMLE+ 2.3 software [23], developed by co-author BR, was used to estimate the age of *BRCA1* c.3331_3334delCAAG. The DMLE+ 2.3 algorithm exploits an intra-allelic coalescent model to assess the linkage

disequilibrium across the marker set coupled to marker locations, population growth rates, and an estimate for the proportion of the disease-bearing chromosomes. For mutation dating analyses, we focused these analyses in Colombia and Iberia as we had the highest number of available carriers and controls from these regions. For Colombia, 28 *BRCA1* c.3331_3334delCAAG carriers and 265 region-matched controls (from Neiva, where the mutation reaches its highest frequency) were used for mutation dating. From Iberia, all Spanish and Portuguese mutation carriers ($n = 15$) and 162 IBS controls (from 1000 Genomes [24]) were used for mutation dating in the peninsula. The population growth rate was estimated as previously reported in Colombia and other parts of the world [25, 26]. Map distances were estimated on the basis of physical distances given by the genetic map HapMap Phase 3.

Colombia is the country with the highest prevalence of the *BRCA1* c.3331_3334delCAAG mutation (~3%) in unselected breast cancer cases [6, 8, 15], and considering the breast cancer incidence, the proportion of mutation-carrying chromosomes is estimated. The proportion of mutation-carrying chromosomes sampled from Colombia was estimated to be a minimum of $f = 0.000012$ (assuming an overall prevalence of *BRCA1* carriers of 0.045) and a maximum of $f = 0.00056$ (assuming an overall prevalence of *BRCA1* carriers of 0.001). Given the prevalence of *BRCA1* carriers of about 1:1000 in the general population and using 46 million as the population of Spain, the proportion of mutation-carrying chromosomes was estimated as $f = 0.00026$ for Spain [27].

Growth rate by generation was estimated with the following equation:

$$\text{Growth rate}_{\text{gen}} = \frac{\ln(P_t/P_o)}{g}$$

where P_t is the current population size, P_o is the initial population size, and g is the number of generations between the current population size and the population size at the moment of mutation origin. The current population size of Colombia is 51 million. Assuming 521 years since the Spanish arrival and 20 years per generation gives $521/20 = 26.05$ generations. Assuming 1000 founders $(51 \times 10^6/1000)/(26.05) = 0.42$ and assuming 100 founders $(51 \times 10^6/100)/(26.05) = 0.51$. We performed mutation age estimates using both values. The generation growth rate of the Spanish population was assumed to be between $d = 0.08$ and 0.11. Results were determined using 100,000 burn-in iterations with 1,000,000 iterations in total for both Colombia and Spain. Additional details of all mutation dating calculations are shown in the [supplementary materials](#).

Genetic ancestry estimation

Global ancestry

Global admixture was performed using Admixture supervised algorithm [28] bootstrapped 200 times and utilized a dataset composed of 1000Genomes super populations (Africans, American, European, East Asian, South Asian) combined with an in-house Indigenous American dataset which included Maya, Aymara, Mixtec, Quechua, Tlapanec, and Nahuatl. To ensure that non-admixed individuals were used in the reference dataset for Admixture, Eigenstrat PCA analysis [29] was performed on the reference dataset and individuals were plotted and filtered using 3 principal components. Only individuals clustered and on ancestral axes that displayed no admixture were included in reference datasets for Admixture and RFMix [30]. In addition, Admixture was run unsupervised with $K=2$ to $K=9$ on the reference dataset and global ancestries were validated. Reference individuals from the 1000Genomes superpopulations displaying no admixture were utilized in Admixture and RFMix. Statistical analysis was performed with Student's t test to examine distributional differences between the ancestry of carriers and non-carriers. All values are expressed as mean \pm SD. $P < 0.05$ was considered as statistically significant.

Local ancestry

For local ancestry estimations, samples were phased using SHAPEIT and then local ancestry was calculated using RFMix PopPhased option using same reference panels as above in EM iterations, 2 EM iterations were performed, and minimum node size of 5 was used—as per recommended settings because the number of individuals in reference populations were skewed. Chromosome 17 global ancestry was calculated using Viterbi predictions of ancestry as the sum of midpoint distances between upstream and downstream markers divided by total chromosome length for ancestral predictions. For regional ancestry plots for *BRCA1* mutation carriers, counts of Amerindian, European, and African ancestry were calculated per marker and then divided by the total number of *BRCA1* mutation carriers in the set.

Results

Haplotype analysis and genetic distance

Using BEAGLE and GERMLINE, two main mutation haplotypes were identified among the *BRCA1* mutation carriers from the six countries (Spain, Colombia, Portugal, Angola, Brazil, and Chile). One shared haplotype was 3.9 Mb long (chr17: 39907129-43807063, between markers rs55675201 and Affx-92039463), and the other haplotype was 2.8 Mb long (chr17: 39788384-42624404, between markers rs4076033 and rs4793119). The first haplotype was shared among individuals from Colombia, Angola, Portugal, Brazil, and Spain, while

the latter was shared only between Chile and Spain. Manual inspection of the mutation region via multiple-sequence alignment revealed a conserved haplotype among all mutation carriers, which was likely too small to detect using the BEAGLE or GERMLINE software. This core mutation haplotype, as determined by BEAGLE (chr17: 41223094-41487451), was flanked by Affx-13890652 and rs75854888, creating boundaries of a 264.4-kb conserved region (Fig. 1).

The largest shared mutation haplotypes were identified among individuals from Colombia (26.5 Mb, chr17: 32835986-59366049, between rs75535552 and rs7215706), while the smallest were between carriers from the Iberian Peninsula. This suggests that the mutation first originated in Iberia as the length of the ancestral haplotype around the mutation is inversely correlated with the number of generations since it first appeared. The phylogenetic tree of the haplotypes was consistent with the previous analysis, where two main haplotypes exist among the mutation carriers. The mutation haplotype likely diverged in Spain prior to the mutation migrating to the other countries (Fig. 2).

Portuguese and Brazilian population mutation haplotype

To verify a shared haplotype among additional Portuguese and Brazilian mutation carriers which became available after SNP genotyping was completed, we genotyped these individuals with 15 SNPs surrounding the *BRCA1* c.3331_3334delCAAG mutation (Fig. 3). These mutation carriers harbored a conserved mutation haplotype that spanned from rs2229611 to rs7214920 (Chr17: 41,063,466-45,051,129), indicating a minimum shared haplotype of 3.9 Mb. In the event that recombination may have occurred within this large window between markers, the two closest flanking markers rs2229611 and rs17599948 (Chr17:41,063,466-41,353,410) to the mutation produced a ~290-kb shared window.

Estimating chromosome 17 European ancestry among Colombian mutation carriers

Given that the mutation likely originated from Spain, we hypothesized that Colombian carriers would be on average, more European along chromosome 17, where *BRCA1* locates, than the average Colombian controls. We found that local ancestry among carriers was higher in the *BRCA1* region (Fig. 4a) and that mutation carriers had higher chromosome 17 European ancestry than non-mutation carriers ($P = 0.000116$, Fig. 4b).

Estimation of allele age in Iberia and Colombia

To estimate the date of the mutation, 60 SNPs residing within a 4.35-Mb window around the *BRCA1* c.3331_3334delCAAG mutation were chosen to be used with the DMLE approach. For Colombia, the mutation age

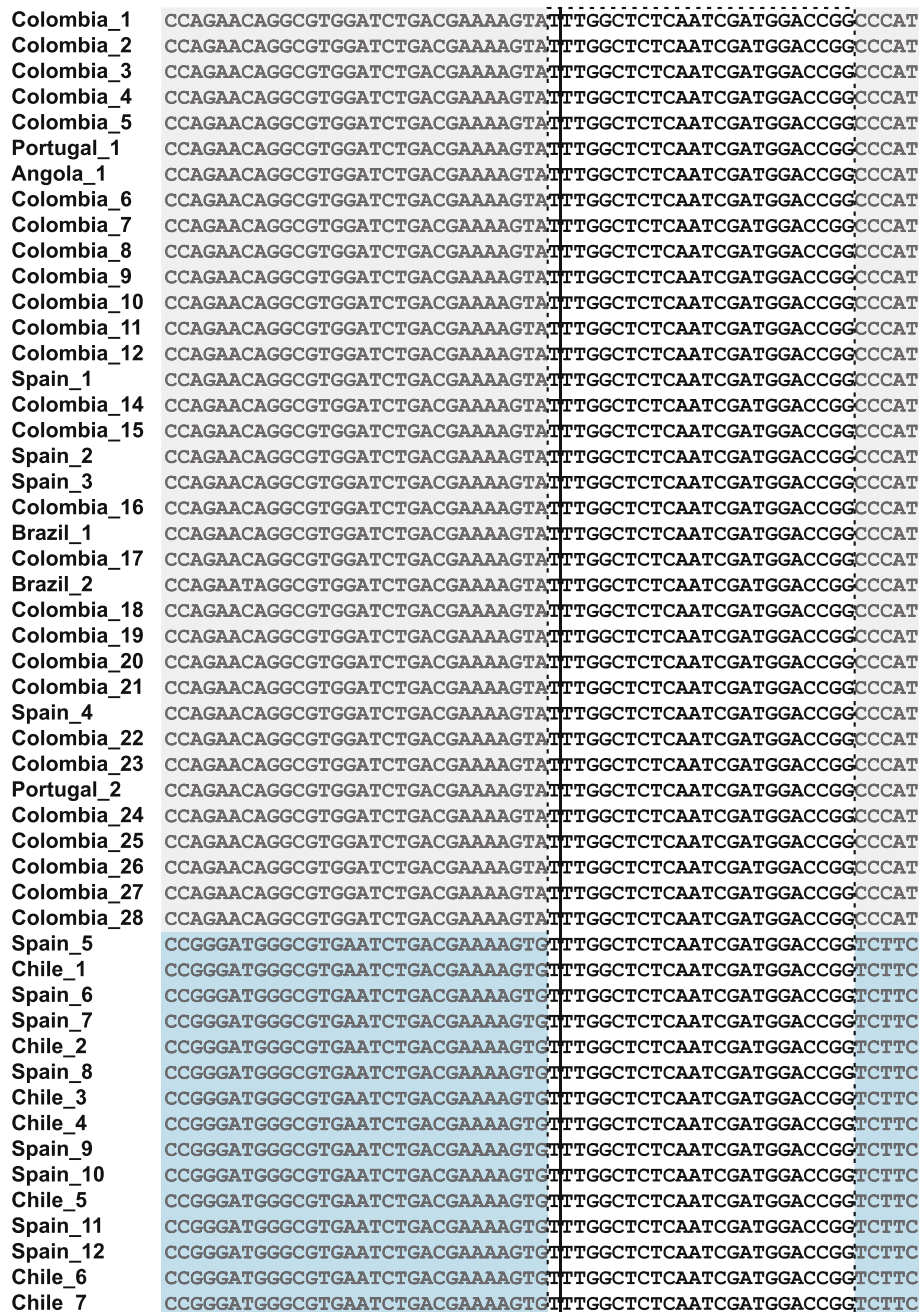


Fig. 1 Multiple-sequence alignment of the mutation haplotype using genome-wide SNP. Data revealed a core haplotype (chr17: 41223094-41487451). The conserved region has a starting marker of Affx-13890652, and ending marker of rs75854888, creating boundaries of a 264.4-kb conserved window (dotted black box) around the mutation (location indicated by solid black line)

estimates in generations (posterior mean and 95% credibility interval) with $f=0.000012$ were 36.3 (31.3, 44.3) assuming $d = 0.42$ and 29.7 (25.4, 36.8) assuming $d = 0.51$. With $f=0.00056$, the estimates were 27.6 (22.5, 36.3) assuming $d = 0.42$ and 24.8 (19.9, 32.3) assuming $d = 0.51$. Assuming 20 years per generation, these mean ages range from 496 to 726 years. For Iberia, using $f=0.00026$, the

mutation age estimates were 121.0 (97.1, 153.6) assuming $d = 0.08$ and 98.0 (75.9, 128.9) assuming $d = 0.11$. Assuming 20 years per generation, these mean ages range from 1960 to 2400 years. These results support the hypothesis that one or a small number of copies of the *BRCA1* mutation were introduced into Colombia via Spanish colonists at the time of the population founding/admixture event.

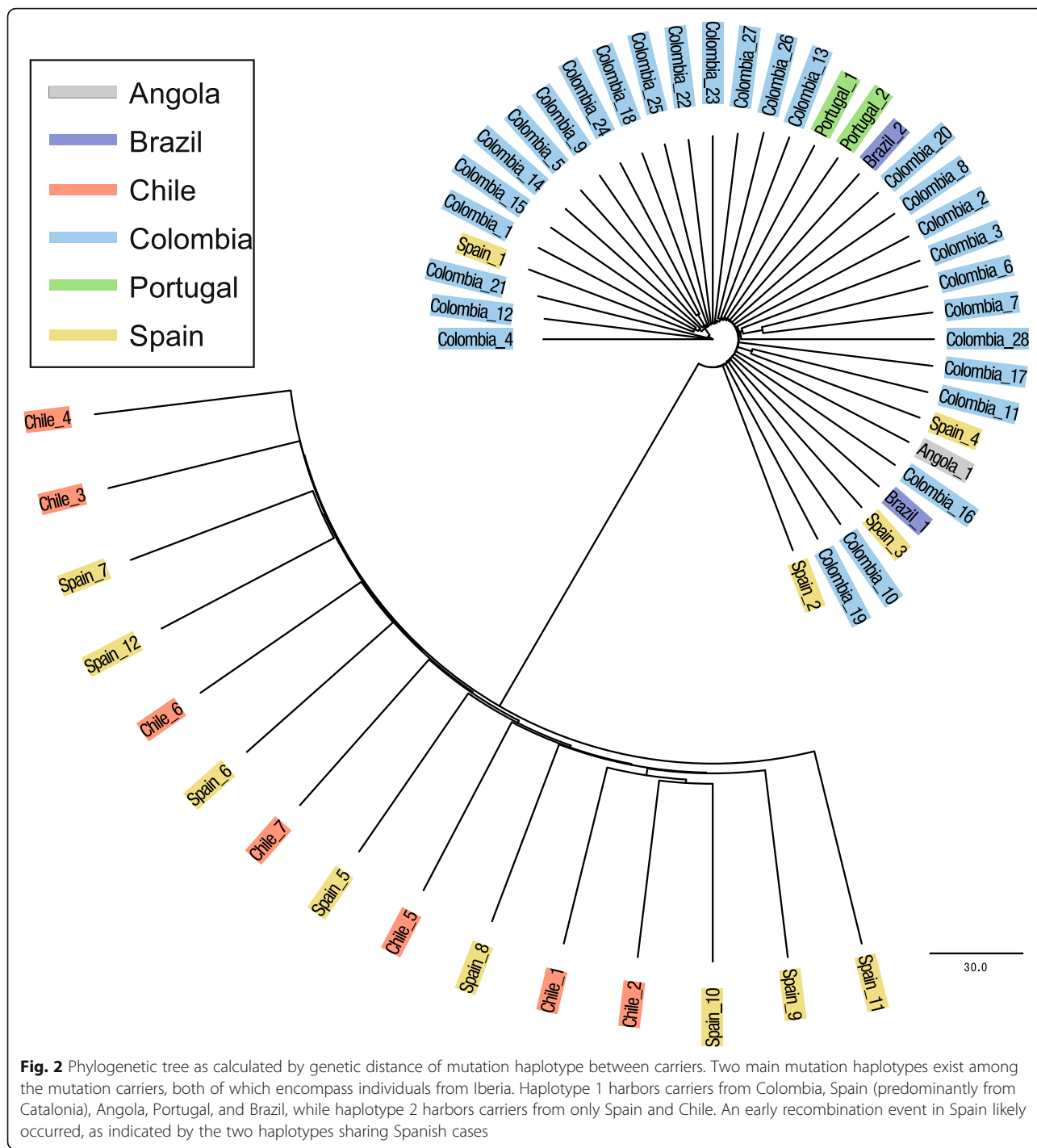


Fig. 2 Phylogenetic tree as calculated by genetic distance of mutation haplotype between carriers. Two main mutation haplotypes exist among the mutation carriers, both of which encompass individuals from Iberia. Haplotype 1 harbors carriers from Colombia, Spain (predominantly from Catalonia), Angola, Portugal, and Brazil, while haplotype 2 harbors carriers from only Spain and Chile. An early recombination event in Spain likely occurred, as indicated by the two haplotypes sharing Spanish cases

Discussion

The comparison of haplotypes between individuals with the same mutation can distinguish whether high-frequency alleles derive from an older or more recent single mutational event and can also determine whether the mutation had arisen independently from multiple individuals. Our study suggests that the *BRCA1* c.3331_3334delCAAG was introduced to Colombia and South

America early in the colonization of the country, resulting in a high mutation prevalence in the population. The estimated age of this mutation in Colombia is consistent with this historical account.

Haplotype length is inversely correlated with the number of generations separating the common ancestor from cases with the mutation in the present time. Our approach revealed a shared mutation haplotype by carriers

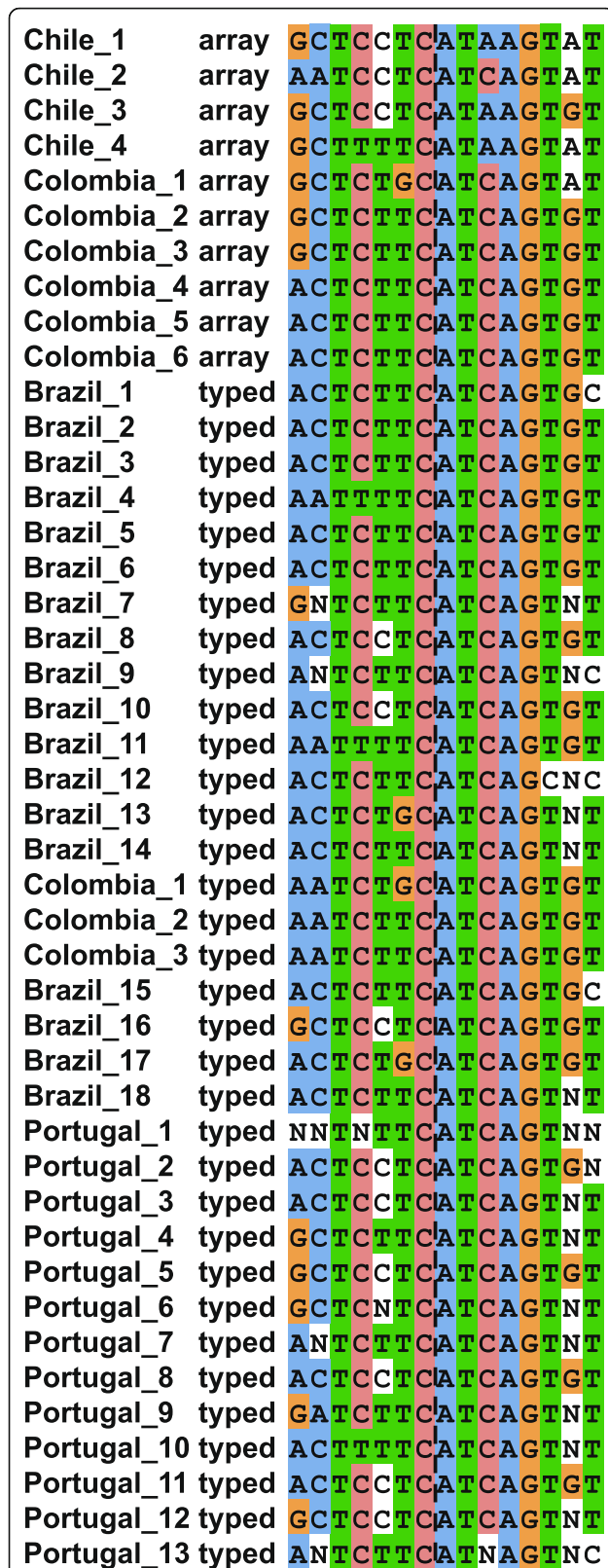


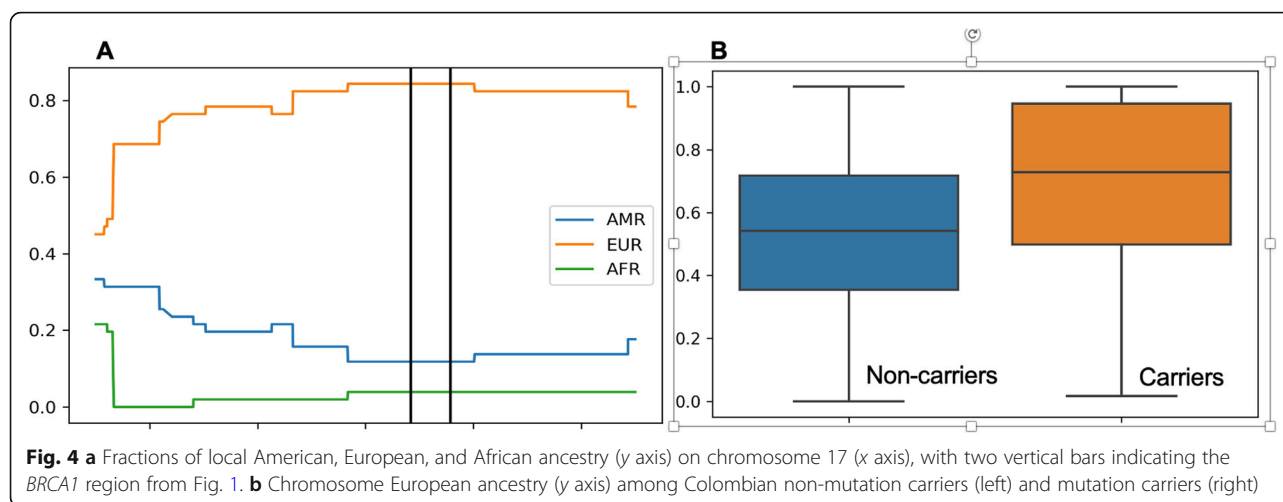
Fig. 3 Haplotypes in 34 *BRCA1* c.3331_3334delCAAG carriers genotyped with 15 flanking SNPs. Black dashed line indicates the location of the mutation

of six countries, multiple continents, and numerous families. These findings depict a history of immigration that is consistent with ancestral links between these populations. The estimated ages from our study and ancestry estimates in Colombian mutation carriers are consistent with the country's history and origin of the mutation, in addition to the genetic demography of Colombia. The mutation was likely introduced to the region during early colonial times during the early 1500s, and our findings in Iberia are consistent with previous dating estimates for other mutations [31]. Moreover, our studies suggest an early recombination event in Spain, which results in the two main haplotypes around the mutation. Spanish and Portuguese colonization of Brazil, Chile, and Colombia during the early 1500s is consistent with the mutation distribution found in our study. In fact, the differences in time periods of Spanish colonization and conquest can be represented by the two main mutation haplotypes found in this study. Interestingly, we also found the same haplotype in a carrier from Angola, a former Portuguese colony, and thus our findings are consistent with the European colonization of Africa and the Americas.

We used genome-wide SNP data to capture the mutation haplotype and estimate mutation age rather than traditional microsatellite markers, which allowed us to comprehensively assess the mutation haplotype via IBD analysis and multiple sequence alignment. A similar approach can be exploited for mapping new variants [32]. We recognize that there may be more to explore surrounding this mutation. While we were able to date the mutation in Iberia and Colombia, we lacked sufficient control data for other countries, such as Chile or Brazil, to allow us to date the mutation in such countries. We anticipate that the mutation age in the other countries will be related to the time of Spanish and Portuguese colonization. We also cannot exclude that the mutation may have multiple ancestral origins in countries without a history of colonization by those countries, such as Canada or Norway, where this mutation has been also reported [7, 33]. Furthermore, while our study in Colombia focused on communities from the central Andean region, where we have shown that they have a predominant European and Indigenous American ancestry [16, 34–40], a recent study in Afro-Colombian populations from the west of the country also identified *BRCA1* c.3331_3334delCAAG carriers, which may suggest additional origins in other Colombian groups [41]. A similar analysis with carriers from these populations would be necessary to confirm this hypothesis.

Conclusions

In summary, we demonstrated the existence of a single ancestral mutation haplotype among six different



countries and general mutation age in the Colombian and Iberian populations are in agreement with historic migration and cultural patterns. Colombian mutation carriers have a higher European ancestry than non-mutation carrier cases, a finding that further support a European origin of *BRCA1* c.3331_3334delCAAG. We also highlight the advantage of utilizing genomic approaches to comprehensively assess founder mutations, since genome-wide SNP data can be exploited to measure ancestry or genetic distance between mutation haplotypes, in addition to haplotype analysis and mutation age estimation.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13058-020-01341-3>.

Additional file 1: Supplementary Table 1. KASP primers used to type mutation haplotype in *BRCA1* c.3331_3334delCAAG mutation carriers from Portugal and Brazil. **Supplementary Table 2.** Summary of mutation carriers and genotype experiments.

Additional file 2.

Abbreviations

BC: Breast cancer; SNP: Single nucleotide polymorphism; IBD: Identical by descent; IBS: Iberian populations in Spain; DMLE+: Bayesian linkage disequilibrium gene mapping

Acknowledgements

We are grateful to all of the individuals who participated in the current study.

COLUMBUS Consortium contributors (in alphabetical order): Fernando Bolaños (Hospital Hernando Moncaleano Perdomo, Neiva, Colombia), Raúl Murillo (Pontificia Universidad Javeriana, Bogotá, Colombia), Yesid Sánchez (Universidad del Tolima, Ibagué, Colombia), Carolina Sanabria (Instituto Nacional de Cancerología, Bogotá, Colombia), Martha Lucia Serrano (Instituto Nacional de Cancerología, Bogotá, Colombia), John Jairo Suarez (Universidad del Tolima, Ibagué, Colombia).

The Brazilian Familial Cancer Network contributors (in alphabetical order); Barbara Alemar (Medical Genomics Laboratory, Hospital de Clinicas de Porto Alegre (HCPA), Porto Alegre, Brazil), Cristina Brinckmann Oliveira Netto (Medical Genetics Service, Porto Alegre, Brazil), Dirce Maria Carraro (Laboratory of Genomics and Molecular Biology, International Research

Center, A.C. Camargo Cancer Center, São Paulo, Brazil; Laboratory of Genomic Diagnostics, Anatomic Pathology Department, A.C. Camargo Cancer Center, São Paulo, Brazil), Fernando Regla Vargas (Birth Defects Epidemiology Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation and Medical Genetics Service, Gaffrée Guinle Hospital, Federal University of Rio de Janeiro State, Rio de Janeiro, Brazil), Gustavo Stumpf da Silva (Medical Genomics Laboratory, Hospital de Clinicas de Porto Alegre (HCPA), Porto Alegre, Brazil), Ivana Lúcia Oliveira Nascimento (Laboratory of Immunology and Molecular Biology (LABIMUNO), Federal University of Bahia (UFBA), Salvador, Bahia, Brazil; Oncology Nucleus of Bahia, NOB, Salvador, Bahia, Brazil), Kelly Rose Lobo de Souza (Genetics Program, National Cancer Institute, Rio de Janeiro, Brazil), Kiyoko Abe-Sandes (Laboratory of Immunology and Molecular Biology (LABIMUNO), Federal University of Bahia (UFBA), Salvador, Bahia, Brazil), Maria Isabel Achatz (Hospital Sírio-Libanês (HSL), São Paulo, São Paulo, Brazil), Miguel Angelo Martins Moreira (Genetics Program, National Cancer Institute, Rio de Janeiro, Brazil), Maria Betânia Torrales (Laboratory of Immunology and Molecular Biology (LABIMUNO), Federal University of Bahia (UFBA), Salvador, Bahia, Brazil), Maristela Pimenta (Laboratory of Genomic Diagnostics, Anatomic Pathology Department, A.C. Camargo Cancer Center, São Paulo, Brazil), Patricia Santos da Silva (Medical Genomics Laboratory, Hospital de Clinicas de Porto Alegre (HCPA), Porto Alegre, Brazil), Taisa Manuela Bonfim Machado-Lopes (Laboratory of Immunology and Molecular Biology (LABIMUNO), Federal University of Bahia (UFBA), Salvador, Bahia, Brazil).

Authors' contributions

AMDAT, PL, JB and LGC-C performed data collection and analyses, wrote and revised the manuscript. BR and BB provided statistical analyses support. JB, CR, AC, AEF, GM, AV, JC, JR, EG, GPE, JS, CA, TT, PAP, AV, CL, ET, CM, MI, MDLH, OD, and COLUMBUS contributed to data acquisition. MB, MRT, PC, ME, and LGC-C designed and supervised the project and secured funding. All authors read and approved the final manuscript.

Funding

AMDEA was supported with graduate Fellowships from UC Davis T32 Biotechnology Program. MB, ME, and LGC-C received funding from GSK Oncology and from "Oficina de Desarrollo a la Docencia de la Universidad del Tolima, Convocatoria 2010". AC and AEF received funding from the Colciencias program "Becas Doctorales Nacionales, Convocatorias 528-2011 and 647-2015". JB received funding from the Colciencias program "Formación de Capital Humano de Alto Nivel para el Departamento de Tolima- 2016, convocatoria 755". *BRCA1/2* genotyping of the Brazilian patients was performed in part with grants from CNPq (408313/2016-1), FAPERGS and Fundo de Incentivo a Pesquisa (FIPE) do Hospital de Clínicas de Porto Alegre, Brazil. AV received funding from CIBERER (grant ER17P1AC7112/2017). PC received funding from FONDEF grant #CA12I10152. CL was Supported by the Carlos III National Health Institute funded by FEDER funds—a way to build Europe—[PI19/00553; PI16/00563; PI16/01898; SAF2015-68016-R and CIBERONC]; the Government of Catalonia [Pla estratègic de recerca i innovació en salut

(PERIS_MedPerCan and URDCat projects), 2017SGR1282 and 2017SGR496]. BR received funding from the National Institutes of Health (R01GM123306). The contents of this article are solely the responsibility of the authors and do not reflect the official views of the National Institutes of Health. LGC-C received from the V Foundation for Cancer Research and from The Auburn Community Cancer Endowed Chair in Basic Science.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Ethics approval and consent to participate

This study was approved by the Ethics Committee at Tolima University; Pontifical Catholic University of Chile; Federal University of Rio Grande do Sul; Fundación Pública Galega de Medicina Xenómica; Institut Català d'Oncologia, Spain; Institute of Genetics and Molecular Biology (UVA-CSIC); Instituto de Investigación Sanitaria San Carlos; Vall d'Hebron Institute of Oncology; University of Washington; Portuguese Oncology Institute of Porto (IPO Porto) and Biomedical Sciences Institute (ICBAS), University of Porto.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial interests.

Author details

¹Genome Center, University of California Davis, Davis, CA, USA. ²Universidad del Tolima, Ibagué, Colombia. ³Hospital Pablo Tobón Uribe, Medellín, Colombia. ⁴Dinámica IPS, Medellín, Colombia. ⁵Hospital Universitario Hernando Moncaleano Perdomo, Neiva, Colombia. ⁶Pontificia Universidad Católica de Chile, Santiago, Chile. ⁷Department of Genetics, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil. ⁸Post-graduate Course in Genetics and Molecular Biology, UFRGS, Porto Alegre, Brazil. ⁹Medical Genetics Service, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil. ¹⁰Fundación Pública Galega de Medicina Xenómica, Grupo de Medicina Xenómica-USC, CIBERER, IDIS, Santiago de Compostela, Spain. ¹¹Hereditary Cancer Program, Catalan Institute of Oncology, Oncobell Program-IDIBELL, Hospital de Llobregat, Barcelona, Spain. ¹²Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain. ¹³Hospital Universitario Cruces, Barakaldo, Spain. ¹⁴Cancer Genetics Group, Institute of Genetics and Molecular Biology (UVA-CSIC), Valladolid, Spain. ¹⁵Laboratorio de Oncología Molecular, Hospital Clínico San Carlos. IdISSC (Instituto de Investigación Sanitaria San Carlos), Madrid, Spain. ¹⁶Grupo de Cáncer Hereditario, Instituto Oncológico Vall d'Hebron (VHIO), Madrid, Spain. ¹⁷Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA, USA. ¹⁸Department of Evolution and Ecology, University of California Davis, Davis, CA, USA. ¹⁹Portuguese Oncology Institute of Porto (IPO Porto) and Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal. ²⁰Division de Investigaciones, Fundación de Genética y Genómica, Ibagué, Colombia. ²¹University of California Davis Comprehensive Cancer Center, Sacramento, CA, USA. ²²Department of Biochemistry and Molecular Medicine, University of California Davis, Sacramento, CA, USA.

Received: 1 April 2020 Accepted: 16 September 2020

Published online: 21 October 2020

References

- Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. *Mol Oncol*. 2010;4:174–91.
- Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, et al. Breast cancer development and progression: risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis*. 2018;5:77–106.
- Nielsen FC, van Overeem HT, Sørensen CS. Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer*. 2016;16:599–612.
- Johannesdóttir G, Gudmundsson J, Bergthorsson JT, Arason A, Agnarsson BA, Eiriksdóttir G, et al. High prevalence of the 999del5 mutation in Icelandic breast and ovarian cancer patients. *Cancer Res*. 1996;56:3663–5.
- Warner E, Foulkes W, Goodwin P, Meschino W, Blondal J, Paterson C, et al. Prevalence and penetrance of BRCA1 and BRCA2 gene mutations in unselected Ashkenazi Jewish women with breast cancer. *J Natl Cancer Inst*. 1999;91:1241–7.
- Torres D, Bermejo JL, Rashid MU, Briceño I, Gil F, Beltran A, et al. Prevalence and penetrance of BRCA1 and BRCA2 germline mutations in Colombian breast cancer patients. *Sci Rep*. 2017;7:4713.
- Durocher F, Tonin P, Shattuck-Eidens D, Kolnick M, Narod SA, Simard J. Mutation analysis of the BRCA1 gene in 23 families with cases of cancer of the breast, ovary, and multiple other sites. *J Med Genet*. 1996;33:814–9.
- Torres D, Rashid MU, Gil F, Umana A, Ramelli G, Robledo JF, et al. High proportion of BRCA1/2 founder mutations in Hispanic breast/ovarian cancer families from Colombia. *Breast Cancer Res Treat*. 2007;103:225–32.
- Rodríguez AO, Llacuachaqui M, Pardo GG, Royer R, Larson G, Weitzel JN, et al. BRCA1 and BRCA2 mutations among ovarian cancer patients from Colombia. *Gynecol Oncol*. 2012;124:236–43.
- Blay P, Santamaría I, Pitiot AS, Luque M, Alvarado MG, Lastra A, et al. Mutational analysis of BRCA1 and BRCA2 in hereditary breast and ovarian cancer families from Asturias (northern Spain). *BMC Cancer*. 2013;13:243.
- Alvarez C, Tapia T, Perez-Moreno E, Gajardo-Meneses P, Ruiz C, Rios M, et al. BRCA1 and BRCA2 founder mutations account for 78% of germline carriers among hereditary breast cancer families in Chile. *Oncotarget*. 2017;8:74233–43.
- Sahasrabudhe R, Lott P, Bohorquez M, Toal T, Estrada AP, Suarez JJ, et al. Germline mutations in PALB2, BRCA1, and RAD51C, which regulate DNA recombination repair, in patients with gastric cancer. *Gastroenterology*. 2017;152:983–986.e6.
- Palmero EI, Carraro DM, Alemar B, Moreira MAM, Ribeiro-dos-Santos Á, Abe-Sandes K, et al. The germline mutational landscape of BRCA1 and BRCA 2 in Brazil. *Sci Rep*. 2018;8:9188.
- Peixoto A, Salgueiro N, Santos C, Varzim G, Rocha P, Soares MJ, et al. BRCA1 and BRCA2 germline mutational spectrum and evidence for genetic anticipation in Portuguese breast/ovarian cancer families. *Familial Cancer*. 2006;5:379–87.
- Benavides J, Suárez J, Estrada A, Bohórquez M, Ramírez C, Olaya J, et al. Breast cancer in six families from Tolima and Huila: BRCA1 3450del4 mutation. *Biomedica*. 2020;40:185–94.
- Fejerman L, Ahmadiyeh N, Hu D, Huntsman S, Beckman KB, Caswell JL, et al. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. *Nat Commun*. 2014;5:5260.
- Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nature Methods*. 2012;9:179–81.
- Browning BL, Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 2013;194:459–71.
- Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. *Am J Hum Genet*. 2011;88:173–82.
- Marroni F, Cipollini G, Peissel B, D'Andrea E, Pensabene M, Radice P, et al. Reconstructing the genealogy of a BRCA1 founder mutation by phylogenetic analysis. *Ann Hum Genet*. 2008;72:310–8.
- Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, et al. Whole population, genome-wide mapping of hidden relatedness. *Genome Res*. 2009;19:318–26.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947–8.
- Reeve JP, Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics*. 2002;18:894–5.
- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
- Aller E, Larrieu L, Jaijo T, Baux D, Espinós C, González-Candelas F, et al. The USH2A c.2299delG mutation: dating its common origin in a southern European population. *European journal of human genetics*. *Nat Publ Group*. 2010;18:788–93.
- Lalli MA, Cox HC, Arcila ML, Cadavid L, Moreno S, Garcia G, et al. Origin of the PSEN1 E280A mutation causing early-onset Alzheimer's disease. *Alzheimers Dement*. 2014;10:S277–S283.e10.
- Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N, et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst*. 1999;91:943–9.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006;38:904–9.

30. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Human Genet.* 2013;93:278–88.
31. Fachal L, Rodríguez-Pazos L, Ginarte M, Toribio J, Salas A, Vega A. Multiple local and recent founder effects of TGM1 in Spanish families. *Plos One.* 2012;7:e33580.
32. G. Puffenberger E, R. Kauffman E, Bolk S, C. Matise T, S. Washington S, Angrist M, et al. Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum Mol Genet* 1994;3:1217–1225.
33. Møller P, Borg A, Heimdal K, Apold J, Vallon-Christersson J, Hovig E, et al. The BRCA1 syndrome and other inherited breast or breast–ovarian cancers in a Norwegian prospective series. *Eur J Cancer.* 2001;37:1027–32.
34. Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, et al. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in Northwest Colombia. *Am J Hum Genet.* 2000;67:1287–95.
35. Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, et al. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet.* 2003;112:534–41.
36. Criollo-Rayó AA, Bohórquez M, Prieto R, Howarth K, Culma C, Carracedo A, et al. Native American gene continuity to the modern admixed population from the Colombian Andes: implication for biomedical, population and forensic studies. *Forensic Sci Int Genet.* 2018;36:e1–7.
37. Bedoya G, Montoya P, García J, Soto I, Bourgeois S, Carvajal L, et al. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a south American population isolate. *PNAS Natl Acad Sci.* 2006;103:7234–9.
38. Hoffman J, Fejerman L, Hu D, Huntsman S, Li M, John EM, et al. Identification of novel common breast cancer risk variants at the 6q25 locus among Latinas. *Breast Cancer Res.* 2019;21:3.
39. Marker KM, Zavala VA, Vidaurre T, Lott PC, Vásquez JN, Casavilca-Zambrano S, et al. Human epidermal growth factor receptor 2-positive breast cancer is associated with indigenous American ancestry in Latin American women. *Cancer Res.* 2020;80:1893–901.
40. Shieh Y, Fejerman L, Lott PC, Marker K, Sawyer SD, Hu D, et al. A polygenic risk score for breast cancer in US Latinas and Latin American women. *J Natl Cancer Inst.* 2020;112:590–8.
41. Vargas E, Lopez DMT, de Deugd R, Gil F, Nova A, Mora L, et al. Low prevalence of the four common Colombian founder mutations in BRCA1 and BRCA2 in early-onset and familial afro-Colombian patients with breast cancer. *Oncologist.* 2019;24:e475–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

