Original Article

Pathogenic and Likely Pathogenic Variants in *PALB2*, *CHEK2*, and Other Known Breast Cancer Susceptibility Genes Among 1054 *BRCA*-Negative Hispanics With Breast Cancer

Jeffrey N. Weitzel, MD [1,2]; Susan L. Neuhausen, PhD[3]; Aaron Adamson, PhD[3]; Shu Tao, PhD[4]; Charité Ricker, MS[2,5]; Asaf Maoz, MD [5]; Margalit Rosenblatt, BA[6]; Bita Nehoray, MS[1,2]; Sharon Sand, BA[1,2]; Linda Steele, MS[3]; Gary Unzeitig, MD[2,7]; Nancy Feldman, MD[2]; Amie M. Blanco, MS[6]; Donglei Hu, PhD[8]; Scott Huntsman, MS[8]; Danielle Castillo, BS[1,2]; Christopher Haiman, ScD[10]; Thomas Slavin, MD[1,2]; and Elad Ziv, MD[6,8,9]

BACKGROUND: Breast cancer (BC) is the most common cancer and related cause of mortality among Hispanics, yet susceptibility has been understudied. *BRCA1* and *BRCA2* (*BRCA*) mutations explain less than one-half of hereditary BC, and the proportion associated with other BC susceptibility genes is unknown. **METHODS:** Germline DNA from 1054 *BRCA*-mutation–negative Hispanic women with hereditary BC (BC diagnosed at age <51 years, bilateral BC, breast and ovarian cancer, or BC diagnosed at ages 51-70 years with ≥2 first-degree or second-degree relatives who had BC diagnosed at age <70 years), 312 local controls, and 887 multiethnic cohort controls was sequenced and analyzed for 12 known and suspected, high-penetrance and moderate-penetrance cancer susceptibility genes (ataxia telangiectasia mutated [*ATM*], breast cancer 1 interacting protein C-terminal helicase 1 [*BRIP1*], cadherin 1 [*CDH1*], checkpoint kinase 2 [*CHEK2*], nibrin [*NBN*], neurofibromatosis type 1 [*NF1*], partner and localizer of BRCA2 [*PALB2*], phosphatase and tensin homolog [*PTEN*], RAD51 paralog 3 [*RAD51C*], *RAD51D*, serine/threonine kinase 11 [*STK11*], and *TP53*). **RESULTS:** Forty-nine (4.6%) pathogenic or likely pathogenic variants (PVs) in 47 of 1054 participants (4.5%), including 21 truncating frameshift, 20 missense, 5 nonsense, and 4 splice variants, were identified in *CHEK2* (n = 20), *PALB2* (n = 18), *ATM* (n = 5), *TP53* (n = 3), *BRIP1* (n = 2), and *CDH1* and *NF1* (both n = 1) and none were identified in *NBN*, *PTEN*, *STK11*, *RAD51C*, or *RAD51D*. Nine participants carried the *PALB2* c.2167_2168del PV (0.85%), and 14 carried the *CHEK2* c.707T>C PV (1.32%). **CONCLUSIONS:** Of 1054 *BRCA*-negative, high-risk Hispanic women, 4.5% carried a PV in a cancer susceptibility gene, increasing understanding of hereditary BC in this population. Recurrent PVs in *PALB2* and *CHEK2* represented 47% (23 of 49) of the total, suggesting a founder effect. Accurate classification of variants was enabled by carefully controlling for ancestry and the increased identification of at-risk Hispanics for screening and prevention. ***Cancer* 2019;0:1-8**. © *2019 American Cancer Society*.

KEYWORDS: breast cancer, checkpoint kinase 2 (CHEK2), disparities, Hispanics, ovarian cancer, partner and localizer of BRCA2 (PALB2), whole-exome sequencing.

INTRODUCTION

Hispanic Americans are one of the fastest growing racial/ethnic groups in the United States, projected to make up 29% of the total population by 2060.[1] One in 12 Hispanic women will be diagnosed with breast cancer (BC), and it is more likely to be at a later, less curable stage than in non-Hispanic women.[2,3] Yet, there is a dearth of Hispanic-specific health care research, particularly in the area of genomic predisposition to BC. Despite clinical studies documenting the efficacy of cancer risk-reduction measures in high-risk individuals, low-income, underinsured, and ethnic minority individuals have a disproportionate burden of cancer and limited access to genetic cancer risk assessment.[4,5]

Family history of BC is a consistent risk factor for BC across different racial/ethnic populations. *BRCA1* and *BRCA2* (*BRCA*) were identified in the 1990s, and mutations in these genes lead to autosomal dominant inheritance of breast and ovarian cancer susceptibility.[6-8] *BRCA* studies in US Hispanics and in Mexican, Colombian, and Peruvian

Corresponding author: Jeffrey N. Weitzel, MD, Division of Clinical Cancer Genomics, City of Hope, 1500 East Duarte Road, Duarte, CA 91010; jweitzel@coh.org

[1]Division of Clinical Cancer Genomics, City of Hope, Duarte, California; [2]Clinical Cancer Genomics Community Research Network, Los Angeles, California; [3]Division of Biomarkers of Early Detection and Prevention, City of Hope, Duarte, California; [4]Integrative Genomics Core, City of Hope, Duarte, California; [5]University of Southern California Norris Comprehensive Cancer Center, Los Angeles, California; [6]University of California at San Francisco Helen Diller Family Comprehensive Cancer Center, San Francisco, California; [7]Gary Unzeitig MD Office, Laredo, Texas; [8]Division of General Internal Medicine, Department of Medicine, University of California at San Francisco, San Francisco, California; [9]University of California at San Francisco Institute for Human Genetics, San Francisco, California; [10]Center for Genetic Epidemiology, Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, California.

See related editorial on pages 1-2, this issue.

The first 2 authors contributed equally to this article.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Additional supporting information may be found in the online version of this article.

DOI: 10.1002/cncr.32083, **Received:** September 21, 2018; **Revised:** December 12, 2018; **Accepted:** January 3, 2019, **Published online** Month 00, 2019 in Wiley Online Library (wileyonlinelibrary.com)

populations indicate a relatively high rate of recurrent pathogenic and likely pathogenic variants (PVs).[9-15] However, PV alleles in *BRCA* account for <50% of hereditary BC and approximately 70% of identified ovarian cancer susceptibility.[16-18]

Since the discovery of *BRCA*, other genes associated with BC susceptibility have been identified.[19] Multigene panels are increasingly being used in the clinical setting; these include genes associated with high (eg, *BRCA1*) and moderate (eg, checkpoint kinase 2 [*CHEK2*]) BC risk. Improved understanding of the spectrum of genetic susceptibility and clarification of gene-specific risks for BC may result in enhanced screening, prevention, and therapeutic strategies for these patients and their families.[20,21] Given that racial/ethnic minority patients are underrepresented in many cohorts, it is particularly important to define risks in diverse populations. We analyzed data for 12 known and suspected high-penetrance and moderate-penetrance BC susceptibility genes in a large cohort of Hispanics with familial BC.

MATERIALS AND METHODS

Study Population

We included women with BC who previously had tested negative for carrying a *BRCA* mutation through clinical gene testing or through the HISPANEL screening test, which incorporates 114 recurrent Hispanic pathogenic mutations.[10] For the *BRCA*-negative women, further inclusion criteria were: age <51 years at BC diagnosis, bilateral BC, BC and ovarian cancer, OR age at diagnosis between 51 and 70 years with a family history of BC in ≥2 first-degree or second-degree relatives diagnosed at age <70 years. These participants were selected from self-identified Hispanic participants from 3 high-risk registry studies, including the City of Hope (COH) Clinical Cancer Genomics Community Research Network,[12,22] the University of California at San Francisco (UCSF) Clinical Genetics and Prevention Program, and the University of Southern California (USC) Norris Comprehensive Cancer Center clinical genetics program. All individuals were consented and enrolled into the study through center-specific institutional review board–approved protocols. We performed whole-exome sequencing (WES) on DNA from 1078 Hispanics with familial BC who met inclusion criteria and 312 Hispanic controls from Southern California.

We also included WES data from 887 participants from the Multiethnic Cohort (MEC) without BC (approximately half had diabetes) who were self-described Hispanics and had undergone WES at the Broad Sequencing Center. These controls are the Hispanic subset of samples from the Exome Aggregation Consortium (ExAC) database.

After sequencing, we excluded 9 BC cases and 5 controls with <20-fold average coverage. We also searched for duplicates and relative pairs using PLINK 1.9 (http://www.cog-genomics.org/plink/1.9/)[23] and removed 2 cases and 2 controls identified as duplicates and 5 cases and 3 controls identified as first-degree relatives. Finally, we excluded 8 cases with previously undetected *BRCA* PVs (negative on HISPANEL). In total, we had 1054 cases and 1189 controls.

Library Construction, Hybridization, and Massively Parallel Sequencing

We used KAPA Hyper Preparation Kits (Kapa Biosystems, Inc) to generate libraries from 500 ng DNA. Each library was assigned a 6-digit DNA barcode sequence and linked to a unique subject identifier. Eight libraries were pooled and hybridized to the SureSelect Clinical Research Exome (Agilent Technologies) kit to capture the exons of all known human transcripts. One hundred base-pair paired-end sequencing on the HiSEQ 2500 Genetic Analyzer (Illumina Inc.) was performed in the COH Integrative Genomics Core (IGC) to an average fold coverage of ×65. Paired-end reads from each sample were aligned to human reference genome (hg37) using the Burrows-Wheeler Alignment Tool (BWA, version 0.7.5a-r405) under default settings, and the aligned binary format sequence (BAM) files were sorted and indexed using SAMtools.[24,25] The same FASTA reference file had been used for aligning the MEC control samples. The sorted and indexed BAMs were processed by Picard MarkDuplicates (version 1.67, http://broadinstitute.github.io/picard/) to remove duplicate sequencing reads. Variant calling from the BAM files from the IGC and the Broad Sequencing Center were processed together at UCSF. After local realignment of reads around in-frame insertions and deletions (indels) and base quality score recalibration by The Genome Analysis Toolkit (GATK, v3.6-0-g89b7209), GATK HaplotypeCaller was used to call variants (https://software.broadinstitute.org/gatk). Variants were considered high confidence if coverage at the site was >10 reads and, for heterozygous calls, if the alternate allele was seen ≥4 times. DNA from 8 MEC participants were sequenced at both COH and the Broad Sequencing Center with >99.8% concordance for variant calling.

Selection of Genes for Analysis

We selected 12 known and suspected, high-penetrance and moderate-penetrance BC and ovarian cancer

susceptibility genes (ataxia telangiectasia mutated [*ATM*], breast cancer 1 interacting protein C-terminal helicase 1 [*BRIP1*], cadherin 1 [*CDH1*], *CHEK2*, nibrin [*NBN*], neurofibromatosis type 1 [*NF1*], partner and localizer of BRCA2 [*PALB2*], phosphatase and tensin homolog [*PTEN*], RAD51 paralog 3 [*RAD51C*], *RAD51D*, serine/threonine kinase 11 [*STK11*], and *TP53*) for this study, based in part on inclusion in clinical multigene panels and as actionable genes in the National Comprehensive Cancer Network guidelines.[26]

Quality Filtering

Samples with average coverage <20-fold across the target region were filtered out (N = 9). Variants with a call quality <20, a read depth <10, a less frequent allele depth of <4, or an allele fraction ratio <30% were filtered out for low quality.

Variant Filtering

Variants with a frequency >2% in the 1000 Genomes Project; the National Heart, Lung, and Blood Institute (NHLBI); the Exome Sequencing Project (ESP); or ExAC databases were removed. Variant call format files were evaluated using Ingenuity Variant Analysis (IVA) version 4 (Qiagen Inc). IVA used the following content versions: Ingenuity Knowledge Base (Hogwarts 160211.000), 1000 Genome Frequency (version 5b),[27] Exome Variant Server (ESP6500SI-V2),[28] ExAC[29] data set (release 0.3), PhyloP,[30,31] Sorting Intolerant From Tolerant (SIFT),[32] the Human Gene Mutation Database (HGMD, version 2015.4), COSMIC (version 75),[33] and ClinVar.[34] American College of Medical Genetics and Genomics (ACMGG) guidelines were applied to the variants using the IVA ACMGG calling algorithm.[35] IVA categorizes variants based on standard ACMGG variant calling recommendations in addition to running in silico models as described above. All ACMGG-called or likely PVs, as well as the remaining frameshift variants, stop codon changes, or variants that disrupt a splice site up to 2 bases into the intron, were individually evaluated by the research team using the available literature and ClinVar to make a final call.[34]

Ancestry Estimation

The Clinical Research Exome included a custom panel of 180 ancestry-informative single nucleotide polymorphisms. On the basis of a previous publication,[36] these markers were selected to be informative for ancestry in a mixed European, Native American, and African population. In addition, we selected 7691 variants common to our WES data and a data set of Axiom arrays, including African (N = 90), European (N = 90), and Native American (N = 51) populations. We selected unlinked markers by linkage disequilibrium pruning in PLINK, identifying a subset of 4544 variants for ancestry estimation. We estimated genetic ancestry using ADMIXTURE and performed analyses with both supervised (specifying the ancestral populations) and unsupervised (including the data from ancestral populations, but not specifying the identity of ancestral populations) runs.

To determine genetic ancestry in the ExAC data, we used the same ancestral reference samples and selected a subset of independent variants (n = 12,758) that overlapped between the Axiom arrays and the ExAC data set. We then entered the allele frequency of these markers into a likelihood function assuming that the ancestry is a 3-component model (African, European, and Native American ancestry). Assuming the markers are independent, the allele frequency for all of the variants in ExAC is a multinomial likelihood function that can be maximized for African, European, and Native American components.

Association Analysis

We performed 2 types of association tests. First, we performed association analyses using ExAC controls, excluding samples from The Cancer Genome Atlas and selected the subset of cases that matched the global ancestry of the ExAC Hispanics. The ExAC samples had been selected to eliminate those with lowest Native American ancestry; therefore, we removed cases with the highest European ancestry until the mean European and Native American ancestry were <1% different from ExAC. The average genetic ancestry in the 547 remaining cases was 57.7% Native American, 7.0% African, and 35.3% European. The average genetic ancestry in the ExAC was 58.7% Native American, 5.4% African, and 35.9% European. We then used the Fisher exact test to compare the allele frequency in cases with ExAC controls. We only performed the comparisons for PVs observed ≥3 times in cases and at least once in the ExAC data set. We calculated odds ratios (ORs) under the assumption that each of the carriers of the variants we tested in the ExAC data set was heterozygous, a reasonable assumption for variants with a frequency <0.5% in the population.

To validate our results, we also performed an analysis using individual data on the 1054 cases and 1189 controls for whom we had individual sequence data. This analysis used the jointly called genotype results, so that the informatics pipeline was identical. To adjust these analyses for genetic ancestry, we added genetic ancestry

as a covariate into logistic regression models rather than matching. To calculate 95% CIs and *P* values for these low frequency variants, we used bootstrap sampling, sampling 10,000 times. All analyses were conducted in R, 3.4.2 (https://www.R-project.org) (The R Project for Statistical Computing).

RESULTS

The demographic and clinical features of the study population are summarized in Table 1. Among 1054 high-risk, *BRCA*-negative Hispanics, 49 (4.6%) PVs were identified in 7 of the 12 known or suspected BC susceptibility genes including: *CHEK2* (n = 20), *PALB2* (n = 18), *ATM* (n = 5), *BRIP1* (n = 2), *TP53* (n = 3), *CDH1* (n = 1), and *NF1* (n = 1) (Table 2). No PVs were observed in *NBN*, *PTEN*, *RAD51C, RAD51D*, or *STK11*. In Supporting Table 1, the distribution of the 25 unique PVs identified in the cases by type of mutation are shown. In Supporting Table 2, the individual PVs found in both cases and controls are shown with the ACMGG criteria[37] for the call as well as information on age at diagnosis, family history (yes/no), and estrogen receptor, progesterone receptor, and human epidermal growth factor 2 status, if available. In Supporting Table 3, all frameshift, splice, nonsense, and nonsynonymous variants observed in cases and controls and the ANNOVAR (Version 2015 Jun 16) output are shown.[38] Of 405 variants, 143 (35%) were not noted previously in ClinVar.

PVs were detected in 47 (4.4%) BC cases; 3 of 47 carried 2 PVs (Table 2). The ages at diagnosis and tumor characteristics were not significantly different between the 47 participants with PVs compared with the 1007 without any PVs; the median age at diagnosis was 42 years (range, 18-70 years) overall and 43 years (range, 26-68 years) for women who were identified as PV carriers. A family history of BC was reported for 51.1% of the PV carriers compared with 37.8% of noncarriers ($P = .074$).

Of the 49 PVs identified in 47 cancer cases (Table 2), 25 were distinct (see Supporting Tables 1 and 2). Among the 25 distinct PVs, 11 were frameshift variants, leading to truncation; 6 were missense variants; 4 were nonsense variants; and 4 were splice site variants. Among controls, there were a total of 18 PVs, of which 8 (44%) were in *ATM*, and 4 were the recurrent mismatch variant (see Supporting Table 2). Three of the 25 PVs occurred in ≥3 individuals (Table 3) and included 2 recurrent frameshift variants in *PALB2* and 1 recurrent missense variant in *CHEK2*. These 3 variants were significantly more commonly observed in the BC cases than in the ExAC Hispanic controls (Table 3).

TABLE 1. Characteristics of the 1054 Hispanic Breast Cancer Cases

Characteristic	Median [Range] or No. (%)
Personal history of BC	
Age at DX of first BC, y	42 [18-70]
Age at DX of second BC, y	49 [28-74]
No. with second BC	98 (9.3)
No. with OC	12 (1.1)
No. with other cancer	54 (5.1)
Family history of BC/OC, n = 933	
None	574 (61.5)
1 FD/SD relative with BC	237 (25.4)
2 FD/SD relative with BC	92 (9.9)
≥3 FD/SD relative with BC	30 (3.2)
OC	94 (10.1)
Unknown family history	121
Tumor ER status, n = 840	
Positive	597 (71.1)
Negative	235 (28.0)
Indeterminate/borderline	8 (1.0)
Unknown	214
Tumor PR status, n = 819	
Positive	504 (61.5)
Negative	305 (37.2)
Indeterminate	10 (1.2)
Unknown	235
Tumor HER2 status, n = 584	
Positive	147 (25.2)
Negative	428 (73.3)
Indeterminate/inconclusive	9 (1.5)
Unknown	470

Abbreviations: BC, breast cancer; DX, diagnosis; ER, estrogen receptor; FD/SD, first-degree/second-degree; HER2, human epidermal growth factor receptor 2 amplification; OC, ovarian cancer; PR, progesterone receptor.

The recurrent *PALB2* c.2167_2168del: p.M723fs was detected in 9 individuals. In an ancestry-matched comparison with ExAC controls, the variant was associated with an OR of 12.9 (95% CI, 3.5-51.2) (Table 3). In analyses of individual data from cases and controls (Table 3), we also observed a highly significant association with this variant ($P < .0001$). Carriers of this variant were significantly more likely to have been diagnosed with ovarian cancer compared with cases who did not carry this variant (see Supporting Table 4) ($P = .004$). *PALB2* c.2411_2412del: p. S804fs, which was detected in 3 individuals, was associated with an OR of 27.5 (95% CI, 2.1-1431.2) (Table 3). If both recurrent *PALB2* mutations are combined, the OR for BC was 13.9 (95% CI, 4.4-47.7). The *CHEK2* PV (c.707T>C: p.L236P) was detected in 14 cases. In an ancestry-matched comparison with the ExAC database values for Hispanics, the OR for BC risk was 3.2 (95% CI, 1.5-6.5; $P = .002$) (Table 3). In the gnomAD,[29] the allele frequency of L236P is approximately 0.002 in Latinos and zero in all other populations, consistent with a founder variant, likely of Native American ancestry.

TABLE 2. The Number of Women With Pathogenic and Likely Pathogenic Variants in Breast Cancer Susceptibility Genes in 1054 Hispanic Patients With Breast Cancer

Gene	Frameshift	Nonsense	Missense	Splicing	Total Variants
ATM	3[a]	1		1[b]	5
BRIP1	2[a]				2
CDH1				1	1
CHEK2	1	1	16	1[c]	20
NF1	1				1
PALB2	14[b]	3		1	18
TP53			3		3
Total	21	5	20	4	49 (4.6%)

Abbreviations: *ATM*, ataxia telangiectasia mutated; *BRIP1*, breast cancer 1 interacting protein C-terminal helicase 1; *CDH1*, cadherin 1; *CHEK2*, checkpoint kinase 2; *NF1*, neurofibromatosis type 1; *PALB2*, partner and localizer of BRCA2; PMS2, postmeiotic segregation increased 2.
[a]One individual had frameshift variants in both *ATM* and *BRIP1*.
[b]One individual had a *PALB2* frameshift variant and an *ATM* splicing variant.
[c]One individual had both a *CHEK2* missense variant and a splicing variant.

TABLE 3. Frequency of Pathogenic and Likely Pathogenic Variants Observed in ≥3 Individuals Compared With the Exome Aggregation Consortium Database and the Current Sequenced Controls

		Analysis Using ExAC Controls			
Gene	Variant	No. of Cases With Variant/ Total No. Ancestry-Matched Cases (%)[a]	No. of ExAC Controls With Variant/Total No. Controls in ExAC (%)[b,c]	OR [95% CI]	*P*
CHEK2	c.707T>C: pL236P	12/ 612 (1.96)	35/5603 (0.63)	3.2 [1.5-6.5]	.0016
PALB2	c.2167_2168del: p.M723fs	9/612 (1.14)	5/5608 (0.09)	12.9 [3.5-51.2]	.00005
PALB2	c.2411_2412del: p. S804fs	3/612 (0.49)	1/5601 (0.02)	27.5 [2.1-1431.2]	.0035
		Analysis Using Individually Sequenced Controls From the City of Hope and the Multiethnic Cohort			
Gene	Variant	No. of Cases With Variant/ Total No. (%)[d]	No. of Controls With Variant/ Total No. (%)	OR [95% CI]	*P*
CHEK2	c.707T>C: pL236P	14/1045 (1.34)	4/1189 (0.34)	4.1 [1.5-22.0]	.039
PALB2	c.2167_2168del: p.M723fs	9/1045 (0.86)	0/1189 (0)		<.0001
PALB2	c.2411_2412del: p. S804fs	3/1045 (0.29)	1/1189 (0.08)	3.7 [0.0, >100.0]	1.0

Abbreviations: *CHEK2*, checkpoint kinase 2; del, deletion; ExAC, Exome Aggregation Consortium; fs, frameshift; OR, odds ratio; *PALB2*, partner and localizer of BRCA2.
[a]The number of cases has been reduced to match the ancestry of cases with that of controls from the ExAC database.
[b]The carrier frequency was calculated based on the assumption that all of the alleles for these variants in the ExAC database are identified in heterozygous carriers.
[c]This variant is at genome position 22:29107982A/G (Build 37) and corresponds to reference single nucleotide polymorphism rs587782471. In the ExAC and gnomAD databases, it is annotated as Leu279Pro.
[d]Only 1045 of the 1054 cases were included in the joint calling.

Analysis of individual data in cases and controls, adjusting for ancestry, found a similar and significant estimate for an association with BC (OR, 4.1; 95% CI, 1.5-22.0; $P = .039$) (Table 3). All *CHEK2* c.707T>C carriers had estrogen receptor–positive BC (see Supporting Table 4), which was significantly higher compared with women who were not identified as carriers of this variant ($P = .024$).

DISCUSSION

This is the first study of Hispanic women with BC that reports on the spectrum and frequency of PVs in BC susceptibility genes beyond *BRCA*. Similar to our previously reported observation of recurrent *BRCA* variants among US Hispanics[12,15,39] and Mexican patients with BC and ovarian cancer,[9,10] we found that 3 recurrent PVs represented 52% of the total. One recurrent *PALB2* PV was responsible for 50% (9 of 18) of all *PALB2* PVs. We previously demonstrated that Hispanics with this *PALB2* variant (c.2167_2168del: p.M723fs) shared a common genotype; Italian BC families with the same variant had a different founder genotype.[40]

Similarly, there was a recurrent *CHEK2* PV, c.707T>C; pL236P, that was observed 14 times (representing 70% of *CHEK2* PV carriers in our study), with an OR of approximately 3.2 (the more conservative point estimate we found.) This variant is listed 5 times as a variant of uncertain significance and once as likely benign in ClinVar (as deposited by 6 commercial

laboratories), although Myriad Genetics distributed revised clinical test reports reclassifying it as likely pathogenic in October of 2017 (J.N.W., written communication, October 2017). Only through a large study focused on Hispanics are we able to report its definitive association with BC, as this variant is rarely seen in individuals of European ancestry. In fact, this variant is only found in Latin American populations in ExAC and not in Africans or African Americans, Europeans, East Asians, or South Asians; this allele frequency is strongly suggestive that the variant originates from indigenous American populations. The estimated carrier frequency of this variant in ExAC Hispanics (excluding The Cancer Genome Atlas) was approximately 0.6%. Therefore, this variant seems comparable in its prevalence and effect size to the *CHEK2* 1100delC mutation in European ancestry populations.[41-44] This association demonstrates the importance of investigating understudied populations as well as the value of multicenter collaborations.

The incidence of clinically actionable PVs in this cohort of high-risk Hispanic BC was 4.7% and was largely driven by the *PALB2* and *CHEK2* founder PVs. The overall prevalence of PVs is similar to other studies, which have reported on the use of panels in high-risk women of mostly European ancestry.[45-47] However, the relative contribution of individual genes in our study is different from the distribution reported in women of European ancestry. When considering the number of unique PVs (see Supporting Table 3), *PALB2* had the highest number. There was a relative paucity of *ATM* PVs (10% of total) compared with published reports in European ancestry populations.[41,42,47]

The relatively large proportions of founder mutations in *PALB2* and *CHEK2* in our study and in *BRCA1* and *BRCA2* found in our previous studies[9-13,15] are indicative of substantial founder effects that contributed to the modern-day Latin American population. These observations are consistent with more systematic population genetic studies of Native American populations,[48,49] which demonstrate a profound bottleneck that occurred with the settling of the Americas.

Our study has several strengths. First, we have sequenced a large number of Hispanic BC cases, adding significantly to what is known about BC susceptibility in this population. Second, we performed analyses carefully controlling for ancestry. Although ExAC is commonly used for rare variant association studies, there are limitations to its use. In particular, ancestry differences between cases and the ExAC controls could bias any associations. This is particularly true in Hispanics, who have ancestry that is quite heterogeneous among individuals.[50,51] To overcome this bias, we subsampled our cases to match the ancestry of the ExAC controls. Furthermore, we performed a second analysis of our cases and a smaller data set of controls adjusting for genetic ancestry as a covariate. To reduce any potential differences from sequencing coverage and informatics, we jointly analyzed the individual case and control data using the same informatics pipeline. The 2 analyses provided similar results, suggesting that we were able to remove potential ethnicity differences from our associations. Because 937 of the controls from the second analysis of individual sequence results were included in the ExAC data set, the second analysis is not an independent statistical validation. Rather, it is experimental validation that one can match cases on aggregate ExAC ancestry data for case-control analyses of ORs.

There are limitations to this study. First, we did not investigate large genomic rearrangements; the frequency of their occurrence is not reported for many of these genes, and thus we may be underestimating the heritability accounted for by known BC susceptibility genes. We previously reported that large genomic rearrangements accounted for 12% (22 of 189) of all *BRCA* mutations identified in a study of 746 high-risk Hispanic patients with BC in the United States,[12] a higher proportion than seen in non-Hispanics.[52] Second, the histopathologic data are limited, so that we do not have sufficient statistical power to investigate associations with mutation status.

In conclusion, as there are gaps in the incorporation of genetic cancer risk assessment for women with BC in the United States,[53,54] this study contributes to our understanding of the spectrum of genetic susceptibility and risks for BC in Hispanics; this important information can enable enhanced screening, prevention, and therapeutic strategies for patients and their families. For example, detection of a *PALB2* or *CHEK2* PV would prompt a recommendation for intensified breast screening, including annual contrast-enhanced breast magnetic resonance imaging, starting at ages 30 to 35 years,[55] and increased colon cancer screening in the setting of a *CHEK2* PV. Insights about recurrent variants continue to illuminate historic population genetics for Hispanics while raising the possibility that the presence of frequent founder variants may enable economic genetic screening[9] in Hispanic populations.

FUNDING SUPPORT

This work was supported by the National Cancer Institute (NCI) (R01CA184585). Research reported in this publication included work performed in the City of Hope Integrative Genomics Core supported by the NCI of the National Institutes of Health (NIH) under award P30CA033572. Elad Ziv was supported in part by K24CA169004.

Susan L. Neuhausen is the Morris and Horowitz Families Endowed Professor. Jeffrey N. Weitzel is the Dr. Norman and Melinda Payson Endowed Professor in Medical Oncology. The City of Hope Clinical Cancer Genomics Community Research Network was supported in part by award RC4A153828 (principal investigator, Jeffrey N. Weitzel) from the NCI and the Office of the Director, NIH; by grant RSGT-09-263-01-CCE from the American Cancer Society; the Breast Cancer Research Foundation; and by grant 02-2013-044 from the Avon Foundation. Charité Ricker and Asaf Maoz were supported by the NIH through the University of Southern California Norris Comprehensive Cancer Center Core Grant (P30CA014089), as well as the Anton B. Burg Foundation, the H. Leslie and Elaine S. Hoffman Cancer Research Chair, the Lynne Cohen Foundation, the Avon Breast Cancer Crusade (05-2015-055), and a gift from Daniel and Maryann Fong. The multiethnic cohort was supported by NIH grants CA164973, CA054281, and CA063464. Sequencing of the multiethnic cohort controls was conducted as part of the Slim Initiative for Genomic Medicine, a project funded by the Carlos Slim Health Institute in Mexico. Seventy-six of the controls were from the California Teachers Study; the collection and data were funded through the NIH (R01 CA77398).

CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

AUTHOR CONTRIBUTIONS

Jeffrey N. Weitzel: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, resources, supervision, writing–original draft, and writing–review and editing. **Susan L. Neuhausen:** Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, writing–original draft, and writing–review and editing. **Aaron Adamson:** Conceptualization, data curation, formal analysis, investigation, methodology, validation, writing–original draft, and writing–review and editing. **Shu Tao:** Data curation, formal analysis, validation, and writing–review and editing. **Charité Ricker:** Resources and writing–review and editing. **Asaf Maoz:** Resources and writing–review and editing. **Margalit Rosenblatt:** Data curation, investigation, and writing–review and editing. **Bita Nehoray:** Resources and writing–review and editing. **Sharon Sand:** Project administration, resources, supervision, and writing–review and editing. **Linda Steele:** Data curation, investigation, methodology, project administration, and writing–review and editing. **Gary Unzeitig:** Resources and writing–review and editing. **Nancy Feldman:** Resources and writing–review and editing. **Amie M. Blanco:** Resources and writing–review and editing. **Donglei Hu:** Data curation, formal analysis, methodology, project administration, and writing–review and editing. **Scott Huntsman:** Data curation and writing–review and editing. **Danielle Castillo:** Data curation, resources, and writing–review and editing. **Christopher Haiman:** Resources and writing–review and editing. **Thomas Slavin:** Resources and writing–review and editing. **Elad Ziv:** Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, writing–original draft, and writing–review and editing.

REFERENCES

1. Colby SL, Ortman JM. Projection of the Size and Composition of the US Population: 2014 to 2060. Current Population Reports, P25-1143. Washington, DC: US Census Bureau; 2014.
2. Huerta EE. Cancer statistics for Hispanics, 2003: good news, bad news, and the need for a health system paradigm change. *CA Cancer J Clin*. 2003;53:205-207.
3. Chlebowski RT, Chen Z, Anderson GL, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. *J Natl Cancer Inst*. 2005;97:439-448.
4. Strasser-Weippl K, Chavarri-Guerra Y, Villarreal-Garza C, et al. Progress and remaining challenges for cancer control in Latin America and the Caribbean. *Lancet Oncol*. 2015;16:1405-1438.
5. Chavarri-Guerra Y, Blazer KR, Weitzel JN. Genetic cancer risk assessment for breast cancer in Latin America. *Rev Invest Clin*. 2017;69:94-102.
6. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*. 1990;250:1684-1689.
7. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian susceptibility gene BRCA1. *Science*. 1994;266:66-71.
8. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*. 1994;265:2088-2090.
9. Villarreal-Garza C, Alvarez-Gomez RM, Perez-Plasencia C, et al. Significant clinical impact of recurrent BRCA1 and BRCA2 mutations in Mexico. *Cancer*. 2015;121:372-378.
10. Villarreal-Garza C, Weitzel JN, Llacuachaqui M, et al. The prevalence of BRCA1 and BRCA2 mutations among young Mexican women with triple-negative breast cancer. *Breast Cancer Res Treat*. 2015;150:389-394.
11. Abugattas J, Llacuachaqui M, Allende YS, et al. Prevalence of BRCA1 and BRCA2 mutations in unselected breast cancer patients from Peru. *Clin Genet*. 2015;88:371-375.
12. Weitzel JN, Clague J, Martir-Negron A, et al. Prevalence and type of BRCA mutations in Hispanics undergoing genetic cancer risk assessment in the southwestern United States: a report from the Clinical Cancer Genetics Community Research Network. *J Clin Oncol*. 2013;31:210-216.
13. Rodriguez AO, Llacuachaqui M, Pardo GG, et al. BRCA1 and BRCA2 mutations among ovarian cancer patients from Colombia. *Gynecol Oncol*. 2012;124:236-243.
14. Torres D, Rashid MU, Seidel-Renkert A, Weitzel JN, Briceno I, Hamann U. Absence of the BRCA1 del (exons 9-12) mutation in breast/ovarian cancer families outside of Mexican Hispanics. *Breast Cancer Res Treat*. 2009;117:679-681.
15. Weitzel JN, Lagos V, Blazer KR, et al. Prevalence of BRCA mutations and founder effect in high-risk Hispanic families. *Cancer Epidemiol Biomarkers Prev*. 2005;14:1666-1671.
16. Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell*. 2002;108:171-182.
17. Walsh T, King MC. Ten genes for inherited breast cancer. *Cancer Cell*. 2007;11:103-105.
18. Walsh T, Casadei S, Lee MK, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A*. 2011;108:18032-18037.
19. Easton DF, Pharoah PDP, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med*. 2015;372:2243-2257.
20. National Comprehensive Cancer Network (NCCN). NCCN Guidelines Genetic/Familial High-Risk Assessment: Breast and Ovarian Version 2.2016. Available at: https://www.nccn.org/professionals/physician_gls/f_guidelines.asp. Accessed May 10, 2016.
21. Weitzel JN, Blazer KR, MacDonald DJ, Culver JO, Offit K. Genetics, genomics, and cancer risk assessment: state of the art and future directions in the era of personalized medicine. *CA Cancer J Clin*. 2011;61:327-359.
22. MacDonald DJ, Blazer KR, Weitzel JN. Extending comprehensive cancer center expertise in clinical cancer genetics and genomics to diverse communities: the power of partnership. *J Natl Compr Canc Netw*. 2010;8:615-624.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
25. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079.
26. National Comprehensive Cancer Network (NCCN). NCCN Genetic/Familial High-Risk Assessment: Breast and Ovarian Version 2.2017. Available at: https://www.nccn.org/professionals/physician_gls/recently_updated.asp. Accessed December 12, 2018.

27. 1000 Genomes Project. 2010. Available at: http://www.internationalgenome.org/. Accessed December 12, 2018.
28. National Heart, Lung, and Blood Institute (NHLBI). NHLBI Grand Opportunity Exome Sequencing Project (ESP), 2016. Available at: https://esp.gs.washington.edu/drupal/. Accessed December 12, 2018.
29. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285-291.
30. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15:1034-1050.
31. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20:110-121.
32. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(web server issue):W452-W457.
33. Wellcome Sanger Institute. COSMIC: Catalogue of Somatic Mutations in Cancer. Available at: https://cancer.sanger.ac.uk/cosmic. Accessed July 13, 2016.
34. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(database issue):D980-D985.
35. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405-424.
36. Galanter JM, Fernandez-Lopez JC, Gignoux CR, et al. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLOS Genet*. 2012;8:e1002554.
37. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet Med*. 2008;10:294-300.
38. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res*. 2010;38:e164.
39. Weitzel JN, Buys SS, Sherman WH, et al. Reduced mammographic density with use of a gonadotropin-releasing hormone agonist-based chemoprevention regimen in BRCA1 carriers. *Clin Cancer Res*. 2007;13(2 pt 1):654-658.
40. Catucci I, Casadei S, Ding YC, et al. Haplotype analyses of the c.1027C>T and c.2167_2168delAT recurrent truncating mutations in the breast cancer-predisposing gene PALB2. *Breast Cancer Res Treat*. 2016;160:121-129.
41. Slavin TP, Maxwell KN, Lilyquist J, et al. The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *NPJ Breast Cancer*. 2017;3:22.
42. Couch FJ, Shimelis H, Hu C, et al. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol*. 2017;3:1190-1196.
43. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and tumor subtype-specific breast cancer risk estimates for CHEK2*1100delC carriers. *J Clin Oncol*. 2016;34:2750-2760.
44. Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, Nordestgaard BG. CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol*. 2008;26:542-548.
45. Slavin TP, Neuhausen SL, Nehoray B, et al. The spectrum of genetic variants in hereditary pancreatic cancer includes Fanconi anemia genes. *Fam Cancer*. 2018;17:235-245.
46. Tung N, Lin NU, Kidd J, et al. Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. *J Clin Oncol*. 2016;34:1460-1468.
47. Rosenthal ET, Bernhisel R, Brown K, Kidd J, Manley S. Clinical testing with a panel of 25 genes associated with increased cancer risk results in a significant increase in clinically significant findings across a broad range of cancer histories. *Cancer Genet*. 2017;218-219:58-68.
48. Wall JD, Jiang R, Gignoux C, et al. Genetic variation in Native Americans, inferred from Latino SNP and resequencing data. *Mol Biol Evol*. 2011;28:2231-2237.
49. Reich D, Patterson N, Campbell D, et al. Reconstructing Native American population history. *Nature*. 2012;488:370.
50. Ziv E, John EM, Choudhry S, et al. Genetic ancestry and risk factors for breast cancer among Latinas in the San Francisco Bay Area. *Cancer Epidemiol Biomarkers Prev*. 2006;15:1878-1885.
51. Fejerman L, John EM, Huntsman S, et al. Genetic ancestry and risk of breast cancer among US Latinas. *Cancer Res*. 2008;68:9723-9728.
52. Judkins T, Rosenthal E, Arnell C, et al. Clinical significance of large rearrangements in BRCA1 and BRCA2. *Cancer*. 2012;118:5210-5216.
53. Kurian AW, Li Y, Hamilton AS, et al. Gaps in incorporating germline genetic testing into treatment decision-making for early-stage breast cancer. *J Clin Oncol*. 2017;35:2232-2239.
54. Childers CP, Childers KK, Maggard-Gibbons M, Macinko J. National estimates of genetic testing in women with a history of breast or ovarian cancer. *J Clin Oncol*. 2017;35:3800-3806.
55. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Version 1.2018: Genetic/Familial High-Risk Assessment: Breast and Ovarian. Fort Washington, PA: National Comprehensive Cancer Network; 2018.