Received: 18 November 2021 | Revised: 14 February 2022 | Accepted: 21 February 2022

DOI: 10.1111/cas.15312

Cancer Science WILEY

ORIGINAL ARTICLE

Association between germline pathogenic variants and breast cancer risk in Japanese women: The HERPACC study

Yumiko Kasugai[1,2] | Tomohiro Kohmoto[3,4] | Yukari Taniyama[5] | Yuriko N. Koyanagi[5] | Yoshiaki Usui[5,6] | Madoka Iwase[1] | Isao Oze[1] | Rui Yamaguchi[3] | Hidemi Ito[5] | Issei Imoto[7] | Keitaro Matsuo[1,2]

[1]Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan

[2]Department of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan

[3]Division of Cancer Systems Biology, Aichi Cancer Center Research Institute, Nagoya, Japan

[4]Department of Human Genetics, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan

[5]Division of Cancer Information and Control, Aichi Cancer Center Research Institute, Nagoya, Japan

[6]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[7]Aichi Cancer Center Research Institute, Nagoya, Japan

Correspondence
Keitaro Matsuo, Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Kanokoden Chikusa-ku Nagoya, Aichi, 464-8681, Japan.
Email: kmatsuo@aichi-cc.jp

Funding information
Aichi Cancer Center Joint Research Project on Priority Areas; Grant-in-Aid for the Third Term Comprehensive 10-year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan; AMED, Grant/Award Number: JP15ck0106177 and JP21ck0106553; Cancer BioBank Aichi; Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, Culture and Technology of Japan, Grant/Award Number: 17015018, 221S0001, JP16H06277(CoBiA) and JP18H03045

Abstract

Approximately 5%–10% of breast cancers are hereditary, caused by germline pathogenic variants (GPVs) in breast cancer predisposition genes. To date, most studies of the prevalence of GPVs and risk of breast cancer for each gene based on cases and noncancer controls have been conducted in Europe and the United States, and little information from Japanese populations is available. Furthermore, no studies considered confounding by established environmental factors and single-nucleotide polymorphisms (SNPs) identified in genome-wide association studies (GWAS) together in GPV evaluation. To evaluate the association between GPVs in nine established breast cancer predisposition genes including *BRCA1/2* and breast cancer risk in Japanese women comprehensively, we conducted a case-control study within the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (629 cases and 1153 controls). The associations between GPVs and the risk of breast cancer were assessed by odds ratios (OR) and 95% confidence intervals (CI) using logistic regression models adjusted for potential confounders. A total of 25 GPVs were detected among all cases (4.0%: 95% CI: 2.6–5.9), whereas four individuals carried GPVs in all controls (0.4%). The OR for breast cancer by all GPVs and by GPVs in *BRCA1/2* was 12.2 (4.4–34.0, $p = 1.74E\text{-}06$) and 16.0 (4.2–60.9, $p = 5.03E\text{-}0.5$), respectively. A potential confounding with GPVs was observed for the GWAS-identified SNPs, whereas not for established

Abbreviations: 95% CI, 95% confidence interval; ACCH, Aichi Cancer Center Hospital; ACMG/AMP, American College of Medical Genetics and Genomics and the Association for Molecular Pathology; B/LB, benign–likely benign variants; BMI, body mass index; BRCA1/2, BRCA1 or BRCA2; GPVs, germline pathogenic variants; GWAS, genome-wide association study; HERPACC, Hospital-based Epidemiologic Research Program at Aichi Cancer Center; HGDV, Human Genetic Variation database; HRT, hormone replacement therapy; NCCN, National Society of Genetic Counselors and National Comprehensive Cancer Network; OR, odds ratio; PY, pack years; SNP, single-nucleotide polymorphism; ToMMo, Tohoku Medical Megabank Organization Genome Variation database; VUS, variants of uncertain significance.

environmental risk factors. In conclusion, GPVs increase the risk of breast cancer in Japanese women regardless of environmental factors and GWAS-identified SNPs. Future studies investigating interactions with environment and SNPs are warranted.

KEYWORDS
breast cancer, case-control study, environmental factors, germline pathogenic variants, single-nucleotide polymorphism

1 | INTRODUCTION

Breast cancer is the most common cancer among women worldwide.[1] Approximately 5%–10% of breast cancers are known to be hereditary, caused by germline pathogenic variants (GPVs) of breast cancer predisposition genes. The most common breast cancer predisposition genes are *BRCA1* and *BRCA2*.[2,3] Women with GPVs of breast cancer predisposition genes are more likely to develop breast cancer at a younger age than women without them.[4,5] Guidelines from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) and from the National Society of Genetic Counselors and National Comprehensive Cancer Network (NCCN) provide detailed criteria for genetic counseling and possible genetic testing for hereditary breast cancer.[6,7] Genetic counseling and appropriate testing in accordance with these guidelines, together with preventive intervention and surveillance for GPV carriers, are expected to reduce risk and overall mortality through primary and secondary prevention. It is therefore likely important to propose genetic testing for GPVs to identify genetically high-risk populations.[8]

Most of the evidence for the effect size of GPVs on breast cancer comes from studies conducted in Europe and the United States, and no large-scale studies in Japanese women were reported until the study by Momozawa et al.[9] Given the potential for heterogeneity in the prevalence of GPVs in breast cancer predisposition genes across ethnicities, however, the accumulation of evidence among East Asians is necessary, including Japanese.[10] It is also interesting to consider the possibility that GPVs confound established environmental risk factors such as drinking and obesity[11] as well as common genetic polymorphisms.[12] Most previous studies evaluating environmental factors in conjunction with GPVs investigated the impact of individual environmental factors among carriers and noncarriers separately,[13,14] and evidence from a comprehensive evaluation of impact in unselected populations is scarce.

Here, we conducted a hospital-based case-control study to examine the association between GPVs of nine breast cancer predisposition genes and breast cancer risk in Japanese women.

2 | METHODS

2.1 | Subjects

All subjects were Japanese women selected from among participants of the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC) between January 2001 and December 2005. The framework of HERPACC has been described elsewhere.[15] In the present study, participants were asked about their lifestyle in a questionnaire and all provided blood samples at the first visit to Aichi Cancer Center Hospital (ACCH). Case subjects were 697 female patients with breast cancer and no previous history of cancer (354 premenopausal and 343 postmenopausal) newly diagnosed between January 2001 and November 2005 at ACCH. Control subjects were 1394 females with no previous history of cancer who were individually matched to the respective case by age (±5 years) and menopausal status (708 premenopausal and 686 postmenopausal) in a 1:2 case-control ratio. After excluding subjects who declined consent, 629 cases and 1153 controls were identified as eligible. Finally, 625 cases and 1133 controls remained with available sequence data for analysis (Figures S1 and S2).

Regarding the opt-out process, we contacted all subjects for whom no information on death was available in the medical records of ACCH or the hospital-based cancer registry in Japan up to February 2017 by postal mail, offering to examine germline variants related to hereditary breast cancer, and possible disclosure of genetic information when GPVs in the *BRCA1* or *BRCA2* (*BRCA1/2*) gene were identified, if the subject preferred. We provided a 2-month deadline for acceptance of this offer. Subjects who rejected genetic examination were excluded from the measurement of germline variants, and those who did not receive the mailed items due to address change were excluded from analysis.

2.2 | Lifestyle information and categorization

Each participant was asked at the first visit to ACCH about their current age, height, weight, smoking and drinking amount, history of hormone replacement therapy (HRT) use (yes, no), regular exercise (yes, no), referral pattern to our hospital (patient discretion, recommendation by family or friends, referral from another clinic, secondary screening after primary screening, and others), reproductive factors (age at first birth, menarche and menopause, lactation, and parity), and family history of breast cancer (yes, no) before the development of the symptoms for which they first visited ACCH.

Body mass index (BMI) was calculated as the weight in kilograms divided by the height in meters squared and classified into the three groups of <23.0, 23.0–24.9, and ≥25.0. Information on smoking was obtained in pack years (PY), calculated by multiplying the number of packs of cigarettes smoked per day by the number of years of smoking, and categorized into the three groups of never, PY < 10, and PY ≥ 10 years. Daily alcohol intake (g/day) was used as a measure

of drinking intensity. It was calculated using information on the frequency of alcohol drinking and total amount of pure alcohol consumed in each drinking session and classified into the three groups of 0 g/day, <23 g/day, and ≥23 g/day.

Single-nucleotide polymorphism (SNP)-based risk groups were defined based on the number of risk alleles in seven breast cancer susceptibility variants, namely very low (scores of <3), low (4-5), moderate (6-7), and high (>8), as reported previously.[16]

2.3 | Sequencing and bioinformatics analysis

Genomic DNA samples for sequencing were isolated from participants' blood samples with QIAamp DNA Blood Mini Kit (Qiagen).

We analyzed all coding regions and flanking intronic sequences of the nine established genes causing hereditary breast cancer.[2] Total length of the target region was 85,142 base pairs (bp). A total of 314 custom-designed primer pairs were designed and optimized using the Fluidigm website "D3 design" (Fluidigm), and the first multiplex PCR was performed on the Fluidigm JUNO system with the Advanta NGS Library Prep Reagent Kit-LP and 192.24 IFCs adding barcode indexes (Fluidigm). Amplicons produced from genomic DNA were harvested and purified by AMPure XP (Beckman Coulter) and sequencing adaptors were added by further PCR. Pooled amplicons were harvested and purified by AMPure XP. After quality was assessed using a Bioanalyzer (Agilent Technologies) and concentration was measured using a Quantus fluorometer (Promega), pooled amplicons were diluted to prepare unidirectional libraries for 2 × 150 bp paired-end sequencing on a NEXT Seq 550 (Illumina).

Sequence reads were divided in each individual by barcode index and trimmed with Trimmomatic_v0.32 using default parameters[17] and pTrimmer_v1.3.2[18] and aligned with Burrows-Wheeler Aligner (BWA)_v0.7.17[19] on hg19 human genome reference sequence. After base recalibration, variant calling was performed using Haplotypecaller of GATK-v3.7 (Genome Analysis Toolkit, Broad Institute).[20]

2.4 | Annotation of variants

The clinical significance of each variant was annotated using Annovar[21] containing known clinical significance information from ClinVar (v 20200316)[22] and population data from the 1000 genomes project,[23] ExAC,[24] Tohoku Medical Megabank Organization Genome Variation database (ToMMo) (v4.7kjpn-20190826),[25] and Human Genetic Variation database (HGVD) (v2.1).[26] Before evaluating the clinical significance of each variant, we excluded low-quality data and variants with no uncommon variant (minor allele frequency = 0) in the ToMMo genome variation database. We then extracted variants within the coding exonic portion or 2 bp away from the exon/intron boundary based on the gene annotation released by the Reference Sequence (RefSeq) database (hg19, Table S1). Through these steps, 412 variants were determined as variants for further evaluation of clinical significance using the ACMG/AMP guidelines as well as the pathogenicity assertions registered in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). In "population data" of the evidence framework in the ACMG/AMP guidelines, each variant was determined to meet the PM2 category using population databases particularly using the Japanese databases (ToMMo genome variation database v4.7k and HGVD v2.1). In "computational and predictive data," each variant was determined to meet the PVS1, PS1, PM4, PM5, or PP3 category. In "functional data," each variant was determined to meet the PS3 category. We did not evaluate evidence of "segregation data," "de novo data," "allelic data," "other database," and "other data" in the evidence framework. In addition, variants classified as "pathogenic" or "likely pathogenic" by expert panels in the ClinVar database were considered pathogenic or likely pathogenic, respectively. All annotations for each variant were reviewed by a clinical genetics expert (I.I.). All pathogenic and likely pathogenic variants were confirmed by Sanger sequencing (Table S2).

2.5 | Statistical analysis

Differences in background characteristics between cases and controls were evaluated using chi-square test or Fisher's exact test as appropriate. Associations between the presence of GPVs and risk of developing breast cancer were estimated by odds ratios (ORs) and 95% confidence intervals (CIs) using log-F (1,1) logistic regression, a penalized likelihood method which is applicable to the analysis of small and sparse datasets such as GPVs.[27,28] We also evaluated the association with a conditional logistic regression model and Firth's logistic regression model in addition to the log F (1,1) logistic regression model,[29] and present the results with two other logistic regression models in Tables S3–S6. In multivariable logistic models, model 1 included only age as a continuous variable, while model 2 included age as a continuous variable, alcohol consumption (categories: 0g/day, <23 g/day, and ≥23 g/day), cumulative exposure to cigarette smoking (categories: never, PY < 10, and PY ≥ 10), menopausal status (with menstruation or menopause), BMI (categories: <23.0, 23.0–24.9, and ≥25.0), physical exercise habits (yes or no), pattern of referral to ACCH (categories: patient's discretion, family recommendation, referral from other clinics, and secondary screening after primary screening), and SNP-based risk group (very low, low, moderate, and high).[16] We excluded family history of breast cancer from covariates because it is an intermediate factor between GPVs and breast cancer (a directed acyclic graph made using DAGitty[30] is shown in Figure S3).

Associations between the nonpathological variants and breast cancer risk were assessed with GPVs added as a covariate. Linear trends (p for trend) were tested by assigning ordinal variables in each category of the number of nonpathogenic variants as continuous variables in each logistic regression model.

All statistical analyses were performed using Stata statistical software v15.1 (Stata Corp.). The Log-F(1,1) logistic model was

TABLE 1 Characteristics of the study population

	Case ($N = 629$)	(%)	Control ($N = 1153$)	(%)	p^a
Age (years)					
<40	89	14.1	134	11.6	0.613
40–49	172	27.3	337	29.2	
50–59	206	32.8	377	32.7	
60–69	125	19.9	237	20.6	
70–	37	5.9	68	5.9	
Median age ± SD	52.0 ± 10.8		52.3 ± 10.6		0.641
Menopausal status					
Premenopausal	309	49.1	553	48.0	0.639
Postmenopausal	320	50.9	600	52.0	
Age at menopause (years)					
<50	105	16.7	237	20.6	0.110
≧50	213	33.9	356	30.9	
Premenopause	309	49.1	553	48.0	
Unknown	2	0.3	7	0.6	
Median age ± SD	49.8 ± 4.9		49.3 ± 4.6		0.432
Family history of breast cancer					
No	567	90.1	1081	93.8	0.006
Yes	62	9.9	72	6.2	
Age at menarche (year)					
≦12	194	30.8	364	31.6	0.470
13–14	304	48.3	537	46.6	
≧15	124	19.7	228	19.8	
Unknown	7	1.1	24	2.1	
Median age ± SD	13.3 ± 1.5		13.4 ± 1.6		0.827
Age at first live birth (year)					
-25 years	248	39.4	569	49.3	<0.001
26 years-	289	45.9	424	36.8	
No delivery	88	14.0	146	12.7	
Unknown	4	0.6	14	1.2	
Median age ± SD	26.2 ± 3.3		25.4 ± 3.4		0.487
Hormone replacement therapy					
No	546	86.8	951	82.5	0.141
Yes	77	12.2	185	16.0	
Unknown	6	1.0	17	1.5	
Ethanol intake (g)/day					
0 g/day	463	73.6	867	75.2	0.636
<23 g/day	130	20.7	231	20.0	
≥23 g/day	29	4.6	40	3.5	
Unknown	7	1.1	15	1.3	
Pack-years (PY)					
0	534	84.9	951	82.5	0.164
<10	41	6.5	76	6.6	
≥10	48	7.6	120	10.4	
Unknown	6	1.0	6	0.5	

TABLE 1 (Continued)

	Case ($N = 629$)	(%)	Control ($N = 1153$)	(%)	p^a
BMI (kg/m^2)					
<23.0	374	59.5	739	64.1	0.106
23.0–24.9	121	19.2	194	16.8	
≧25.0	134	21.3	212	18.4	
Unknown	0	0.0	8	0.7	
Median BMI ± SD	22.7 ± 3.3		22.4 ± 3.2		0.060
Regular exercise					
No	372	59.1	706	61.2	0.388
Yes	257	40.9	447	38.8	
Pattern of referral to Aichi Cancer Center					
Patient discretion	167	26.6	350	30.4	0.141
Family recommendation	144	22.9	178	15.4	
Referral from other clinic	185	29.4	232	20.1	
Secondary screening after primary screening	125	19.9	381	33.0	
Other	5	0.8	9	0.8	
Unknown	3	0.5	3	0.3	
SNP-based risk group					
Very low (<3)	80	12.7	219	19.1	<0.001
Low (4–5)	248	39.4	505	44.0	
Moderate (6–7)	227	36.1	351	30.6	
High (>8)	74	11.8	74	6.4	

Abbreviations: BMI, body mass index; SD, standard deviation; SNP, single-nucleotide polymorphism.

[a]Differences between cases and controls were analyzed using the unpaired *t* test and Chi-squared test.

conducted using the "penlogit" command[31] and Firth's logistic model using the "firthlogit" command by Coveney (Ref: https://core.ac.uk/display/19370376). We defined *P*-values less than 0.05 as showing statistical significance.

3 | RESULTS

3.1 | Participant characteristics

Table 1 shows the characteristics of case and control participants. Age and menopausal status were appropriately matched between groups. Family history of breast cancer ($p = 6.00E-03$), later age at first birth ($p < 1.00E-03$), and moderate-to-high SNP-based risk groups ($p < 1.00E-03$) were significantly prevalent among cases, whereas other factors did not significantly differ between the groups.

3.2 | Germline variants of breast cancer predisposition genes in Japanese women

We identified 412 germline variants, including 25 GPVs, in 625 breast cancer cases and 1133 controls (Table S1). A total of 105 of the 412 variants that were not annotated in ClinVar were treated as "variants of uncertain significance (VUS)". Log F (1,1) and Firth's logistic regression models successfully estimated ORs and 95% CIs for low-frequency variants which could not be assessed by the conditional logistic regression model. Because CIs were narrower with the log F (1,1) logistic model than with Firth's models, we decided to present estimates by the log F (1,1) logistic model afterward. The location of GPVs and number of subjects are shown in Figure S4. The number of GVPs was highest in *BRCA2*, and GVPs were located in the whole coding region of this gene.

3.3 | Impact of GPVs of breast cancer predisposition genes on the risk of breast cancer

Table 2 shows the association between GPVs and breast cancer risk as evaluated. All GPVs in nine genes were found in 25 cases (4.0%, 95% CI: 2.6–5.9) and in four noncancer controls (0.4%, 0.1–0.9). Germline pathogenic variants in *BRCA1/2* were most common in cases ($n = 19$, 3.0%, 1.8–4.7) and was also detected in controls ($n = 2$, 0.2%, 0.02–0.6). One case had two GPVs, in one each in *BRCA1* and *BRCA2*. A significant association was observed between all GPVs or GPVs in *BRCA1/2* and breast cancer risk. ORs for breast cancer by all GPVs in model 1 and model 2 were 10.40 (95% CI:

TABLE 2 Association between germline pathogenic variants and breast cancer risk

Gene	No. of pathogenic variants	Case ($n = 625$)		Control ($n = 1133$)			logF(1,1) logistic regression					
							Model 1			Model 2		
		No. of carriers	(%)	No. of carriers	(%)	p[a]	OR[a]	95% CI	*P*-value	OR[b]	95% CI	*P*-value
All pathogenic variants in nine genes	25	25[c]	4.00	4	0.35	<0.001	10.40	(3.80–28.49)	5.24E−06	12.20	(4.38–34.01)	1.74E−06
BRCA1/2	19	19[c]	3.04	2	0.18	<0.001	14.08	(3.76–52.77)	8.76E−05	15.97	(4.18–60.94)	5.03E−05
ATM	2	2	0.32	2	0.18	0.618	1.62	(0.28–9.48)	5.93E−01	2.43	(0.40–14.81)	3.36E−01
BRCA1	6	7	1.12	0	0	0.001	25.32	(1.44–445.29)	2.72E−02	29.59	(1.65–530.67)	2.15E−02
BRCA2	13	13	2.08	2	0.18	<0.001	9.58	(2.47–37.12)	1.08E−03	10.97	(2.76–43.57)	6.61E−04
CDH1	0	0	0	0	0	NE	NE	NE				
CHEK2	1	1	0.16	0	0	0.354	4.24	(0.17–107.27)	3.81E−01	3.34	(0.13–82.70)	4.62E−01
PALB2	1	1	0.16	0	0	0.354	4.18	(0.17–105.65)	3.85E−01	4.84	(0.19–126.21)	3.44E−01
PTEN	1	1	0.16	0	0	0.354	4.22	(0.17–107.23)	3.83E−01	3.86	(0.15–97.19)	4.12E−01
STK11	0	0	0	0	0	NE	NE	NE				
TP53	1	1	0.16	0	0	0.354	3.94	(0.16–99.04)	4.04E−01	3.36	(0.14–83.43)	4.59E−01

Abbreviations: CI, confidence interval; OR, odds ratio.

[a]Fisher's exact test.

[b]Adjusted for age as a continuous variable.

[c]Adjusted for age as a continuous variable, alcohol drinking, cumulative exposure to cigarette smoking, menopausal status, body mass index (BMI) in three categories, physical exercise, pattern of referral to Aichi Cancer Center and single-nucleotide polymorphism (SNP)-based risk group (very low, low, moderate, and high by Sueta et al. 2012 BCRT).

[d]One case had two germline pathogenic variants (GPVs) in BRCA1 and 2.

3.80–28.49, $p = 5.24E-06$) and 12.20 (4.38–34.01, $p = 1.74E-06$), respectively, while those by GPVs in *BRCA1/2* were 14.08 (3.76–52.77, $p = 8.76E-05$) and 15.97 (4.18–60.94 $p = 5.03E-05$), respectively. Sensitivity analysis by adding each potential confounder to model 1 demonstrated that no variables except SNPs and family history of breast cancer were associated with any remarkable change in the estimated OR, indicating a lack of confounding for these variables (Figure S5).

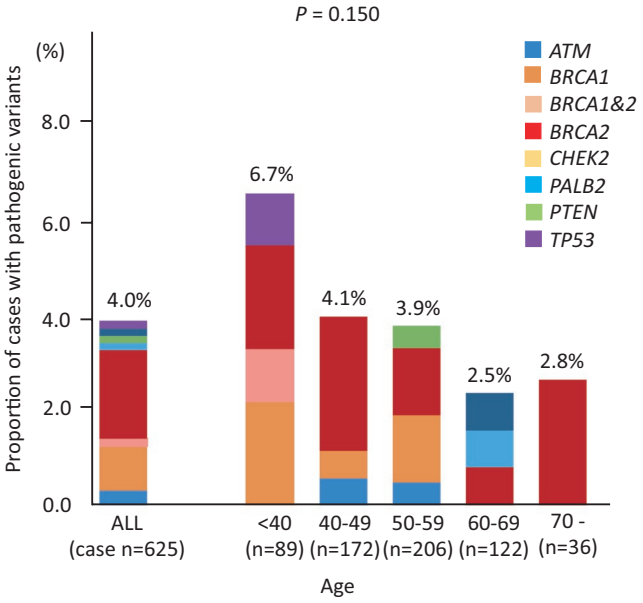

FIGURE 1 Proportion of cases with pathogenic variants decreased with advancing age (nonparametric test for a trend $p = 1.50E-01$). The color code for individual genes is shown in the legend at right

The impact of GVPs by gene is presented in Table 2. We observed significant associations between breast cancer and GPVs in *BRCA1* (OR in model 2 = 29.59, 1.65–530.67, $p = 2.15E-02$) and *BRCA2* (OR in model 2 = 10.97, 2.76–43.57 $p = 6.61E-04$), whereas associations with GPVs in *ATM*, *CHEK2*, *PALB2*, *PTEN*, and *TP53* showed no statistical significance. We observed no GPVs in *CDH1* and *STK11*.

3.4 | Impact of GPVs of breast cancer predisposition genes on the risk of breast cancer by age group

Figure 1 shows the proportion of cases with GPVs by age group. The highest proportion of cases with GPVs in nine genes was 6.7% (2.2% for *BRCA1*, 1.1% for *BRCA1* and *BRCA2*, 2.2% for *BRCA2*, and 1.1% for *TP53*) for patients aged less than 40 years. The proportion of cases with GPVs appeared to decrease with increasing age, although this trend was not statistically significant ($p = 1.50E-01$). The cases with GPVs in nine genes were diagnosed at a younger age (median 48 years) than cases without GPVs (median 52 years; $p = 3.60E-02$).

Table 3 shows age group–stratified associations between GPVs and breast cancer risk. We observed a significant association between all GPVs and breast cancer risk for age groups less than

TABLE 3 Association between germline pathogenic variants and breast cancer risk by age group

			logF(1,1) logistic regression					
			Model 1			Model 2		
Age group	Case ($n = 625$) No. of carriers	Control ($n = 1133$) No. of carriers	OR	95% CI	*P*-value	OR[d]	95% CI	*P*-value
All pathogenic variants in nine genes								
<40	6	0	19.13	(1.06–343.79)	4.52E−02	23.46	(1.22–451.36)	3.65E−02
40–49	7	0	27.88	(1.58–492.02)	2.31E−02	31.30	(1.68–584.69)	2.11E−02
50–59	8	2	6.07	(1.46–25.17)	1.30E−02	6.54	(1.53–27.97)	1.12E−02
60–69	3	2	2.44	(0.47–12.66)	2.87E−01	2.40[a]	(0.44–12.99)	3.09E−01
70-	1	0	4.23	(0.17–107.88)	3.83E−01	8.52[a]	(0.23–314.95)	2.45E−01
Pathogenic variants in BRCA1/2								
<40	5	0	15.87	(0.87–290.55)	6.24E−02	20.52	(1.02–414.15)	4.87E−02
40–49	6	0	23.79	(1.33–425.89)	3.13E−02	21.83	(1.16–411.41)	3.96E−02
50–59	6	2	4.54	(1.04–19.79)	4.38E−02	4.79	(1.05–21.82)	4.27E−02
60–69	1	0	4.27	(0.17–108.34)	3.79E−01	4.63[a]	(0.18–122.57)	3.59E−01
70-	1	0	4.23	(0.17–107.88)	3.83E−01	8.41[a]	(0.24–298.46)	2.42E−01

Abbreviations: CI, confidence interval; OR, odds ratio.

[a]Adjusted for alcohol drinking, cumulative exposure to cigarette smoking, menopausal status, body mass index (BMI) in three categories, physical exercise, pattern of referral to Aichi Cancer Center and single-nucleotide polymorphism (SNP)-based risk group (very low, low, moderate, and high by Sueta et al. 2012 BCRT).

[b]Adjusted for alcohol drinking, cumulative exposure to cigarette smoking, BMI in three categories, physical exercise, reason for referral to Aichi Cancer Center and SNP-based risk group (very low, low, moderate, and high by Sueta et al. 2012 BCRT). Because all participants were menopausal, menopausal status was not included in the covariates.

60 years. Germline pathogenic variants (GPVs) in *BRCA1/2* also showed a significant association with breast cancer with age less than 60 years.

3.5 | Impact of GPVs on breast cancer risk by type of variants

Table 4 shows the association between GPVs and breast cancer risk by type of genetic variant. There were 37 for loss-of-function variants (12 for stop gain, 17 for flameshift short insertion/deletion (in/del), and eight for splicing site), 21 with clinical significance for pathogenic/likely pathogenic, one for conflicting interpretations of pathogenicity, 15 for VUS, and one for benign in ClinVar. Variant types of heterozygous loss-of-function were significantly associated with risk in each model: stop gain/loss (OR in model 2 = 14.86, 95% CI: 2.66–82.92, p = 2.09E-03), flame shift short in/del (5.03, 2.17–11.66, p = 1.66E-04), and splicing site (1.42, 1.14–1.78, p = 2.07E-03).

3.6 | Impact of germline nonpathogenic variants on breast cancer risk

Table 5 shows associations between nonpathogenic variants and breast cancer risk. We categorized non-GPVs that were annotated by ClinVar into two groups, benign–likely benign variants (B/LB) and VUSs. Odds ratios for all non-GPVs and the B/LB group showed no statistically significant association with breast cancer risk. In contrast, we observed a significant association between VUSs and breast cancer risk (OR in model 1 = 1.25, 95% CI: 1.03–1.53, p = 2.52E-02). Moreover, as the number of all non-GPVs and benign variants retained increased, breast cancer risk increased, albeit without statistical significance.

4 | DISCUSSION

In this study, we identified 412 germline variants in nine established breast cancer predisposition genes in 629 breast cancer cases and 1153 controls in Japanese women. Among them, GPVs were identified in 25 breast cancer cases (4.0%) and four noncancer controls (0.4%). The prevalence of cases with GPVs was higher in younger patients, and highest in those aged under 40 years. GPVs in nine genes, including *BRCA1/2*, showed a significantly strong association with breast cancer risk. This association was consistent even after consideration of epidemiologically established environmental factors or SNPs as covariates. These findings reveal that GPVs are powerful risk factors which induce the development of breast cancer at younger ages independently of adjusted factors. This is the first study to comprehensively examine the impact of GPVs on breast cancer risk in conjunction with environmental factors and SNPs in a Japanese population.

TABLE 4 Association between genetic variant type and breast cancer risk

Variant type	No. of variants	Case (n = 625) No. of carriers	Control (n = 1133) No. of carriers	logF(1,1) logistic regression					
				Model 1[b]			Model 2[a]		
				OR	95%CI	P-value	OR	95%CI	P-value
Stop gain/loss	12	12	1	14.75	(2.70–80.57)	1.89E−03	14.86	(2.66–82.92)	2.09E−03
Flameshift in/del	17	19	9	3.70	(1.69–8.10)	1.05E−03	5.03	(2.17–11.66)	1.66E−04
Non-flameshift in/del	5	23	31	1.31	(0.77–2.22)	3.22E−01	1.34	(0.77–2.32)	2.97E−01
Nonsynonymous SNV	263	625	1125	1.01	(0.97–1.06)	5.19E−01	1.00	(0.96–1.05)	9.08E−01
Synonymous SNV	107	622	1128	1.01	(0.96–1.07)	7.05E−01	1.01	(0.96–1.07)	7.19E−01
Splicing site	8	476	790	1.39	(1.12–1.73)	2.68E−03	1.42	(1.14–1.78)	2.07E−03

Abbreviations: CI, confidence interval; in/del, insertion deletion; OR, odds ratio; SNV, single-nucleotide variant.

[a]Adjusted for age as continuous variable.

[b]Adjusted for alcohol drinking, cumulative exposure to cigarette smoking, menopausal status, BMI in three categories, physical exercise, pattern of referral to Aichi Cancer Center and SNP-based risk group (Very low, Low, Moderate, and High by Sueta et al. 2012 BCRT).

TABLE 5 Association between clinical significance and breast cancer risk

Clinical significance in ClinVar	No. of variants	Case ($n = 625$) No. of carriers	Case ($n = 625$) Frequency of carriers	Control ($n = 1133$) No. of carriers	Control ($n = 1133$) Frequency of carriers	logF(1,1) logistic regression						
						Model 1			Model 2			
						OR[b]	95%CI	*P*-value	OR[a]	95%CI	*P*-value	*p* for trend
All nonpathogenic variants	387	625	1.00	1133	1.00							
Subjects with 1–9 nonpathogenic variants		253	0.40	484	0.43	Reference			Reference			
Subjects with ≧ 10 nonpathogenic variants		372	0.60	649	0.57	1.11	(0.91–1.35)	3.16E−01	1.06	(0.86–1.30)	5.83E−01	0.583
All benign	122	625	1.00	1133	1.00							
Subjects with 1–9 benign variants		271	0.43	519	0.46	Reference			Reference			
Subjects with ≧ 10 benign variants		354	0.57	614	0.54	1.11	(0.91–1.35)	2.99E−01	1.07	(0.88–1.32)	4.93E−01	0.493
All uncertain significance	264	325	0.52	533	0.47							
Subjects without VUS (0)		300	0.48	602	0.53	Reference			Reference			
Subjects with 1–5 VUS		325	0.52	531	0.47	1.25	(1.03–1.53)	2.52E−02	1.20	(0.98–1.47)	7.45E−02	0.074

Abbreviations: CI, confidence interval; OR, odds ratio; VUS, variants of uncertain significance.

[a]Adjusted for age as a continuous variable and germline pathogenic variants.

[b]Adjusted for alcohol drinking, cumulative exposure to cigarette smoking, menopausal status, BMI in three categories, physical exercise, pattern of referral to Aichi Cancer Center, SNP-based risk group (very low, low, moderate and high by Sueta et al. 2012 BCRT) and germline pathogenic variants.

In the present study, the prevalence of both GPVs (4.0%) in nine genes and GPVs in *BRCA1/2* (3.0%) in patients with breast cancer was lower than those previously reported from Japan,[9] Europe and the United States,[2–4] and China.[32] This might be partly attributable to two factors: (1) we only accepted GPVs annotated by ClinVar, and (2) the number of targeted genes was smaller than that in previous reports. In addition, the prevalence of GPVs among breast cancer patients can vary across ethnicities and populations. The point estimates of ORs for GPVs by genes were relatively but not significantly lower than those in a previous study from Japan[9] (Table S7). This difference can be explained by the difference in control subject selection: the former study[9] excluded individuals with a family history of cancer, while the present study did not. According to ClinVar, moreover, seven GPVs (1.2%), namely two for *BRCA1*, two for *BRCA2*, one for *CHEK2*, one for *PALB2*, and one for *PTEN*, were novel,[32–40] indicating the potential importance of further accumulation of data in East Asian populations.

As expected, the prevalence of GPVs was higher in younger patients than older patients. Of note, the association was significant among those aged less than 60 years. Current NCCN guidelines (v2.2021) indicate that breast cancer patients aged under 50 years should be provided personalized risk assessment, genetic counseling, genetic testing, and management. A recent report of a case series from Japan demonstrated that a quarter of the cases with GPVs did not meet the NCCN criteria for assessment as high-risk for genetic or familial cancers in a previous Japanese study.[41] Accordingly, the target population for assessment of hereditary breast cancer in Japanese women may differ from that in other populations.

In the analysis by variant type, breast cancer risk was significantly high for structural variants classified as GPVs and VUS in ClinVar. The nine genes we investigated have been reported to cause breast cancer due to loss of function.[42–50] Therefore, it is possible that some pathogenic variants lacking function completely (loss of function) or partially (hypomorphic) are currently classified as VUS due to a lack of functional or genetic data. Indeed, VUSs contain a mixture of variants that cause protein loss of function and nonsense mutations. Some VUS in the ClinVar databases will likely be classified as pathogenic in future.

This study has several methodological strengths. First, because controls were selected from the same population, case-control subjects are assumed to be comparable, warranting the internal validity of this case-control study. Second, we considered potential confounding by individual matching of cases by age and menopausal status, as well as by statistical adjustments in the models. As presented in Figure S5, we speculate that some level of confounding occurs only with SNP risk groups. Although some residual confounding might remain, the association between GPVs and breast cancer risk appears sufficiently strong. The major limitation of this study is a lack of breast cancer subtype–specific analysis due to the limited number of subjects. Further analyses using larger sets are needed to clarify the association between GPVs and individual subtypes of breast cancer.

In conclusion, we confirmed the strong association of GPVs with breast cancer risk in a study which considered potential confounding due to environmental factors and genome-wide association study (GWAS)-identified SNPs in a Japanese population. Moreover, the association was significant in patients aged under 60 years, in contrast with recent NCCN guidelines recommending further genetic testing at under 50 years old, suggesting that a Japanese population–specific algorithm for screening of patients with hereditary breast cancer may be required. Further studies are warranted, particularly in East Asian and other less-investigated populations.

ACKNOWLEDGEMENT

This study is supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, Culture and Technology of Japan (17015018, 221S0001, JP16H06277[CoBiA], JP18H03045), a Grant-in-Aid for the Third Term Comprehensive 10-year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan, AMED (JP15ck0106177, JP21ck0106553), Aichi Cancer Center Joint Research Project on Priority Areas, and Cancer BioBank Aichi.

DISCLOSURE

The authors have no conflict of interest.

ORCID

Yuriko N. Koyanagi https://orcid.org/0000-0002-5675-3429
Isao Oze https://orcid.org/0000-0002-0762-1147
Hidemi Ito https://orcid.org/0000-0002-8023-4581
Keitaro Matsuo https://orcid.org/0000-0003-1761-6314

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
2. 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.
3. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet*. 2002;31:33-36.
4. 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.
5. Couch FJ, Shimelis H, Hu C, et al. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol*. 2017;3:1190-1196.
6. Hampel H, Bennett RL, Buchanan A, Pearlman R, Wiesner GL. A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: referral indications for cancer predisposition assessment. *Genet Med*. 2015;17:70-87.
7. Daly MB, Pal T, Berry MP, et al. Genetic/familial high-risk assessment: breast, ovarian, and pancreatic, version 2.2021, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2021;19:77-102.

8. Katanoda K, Hori M, Matsuda T, et al. An updated report on the trends in cancer incidence and mortality in Japan, 1958–2013. *Jpn J Clin Oncol*. 2015;45:390-401.
9. Momozawa Y, Iwasaki Y, Parsons MT, et al. Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls. *Nat Commun*. 2018;9:4083.
10. Kaneyasu T, Mori S, Yamauchi H, et al. Prevalence of disease-causing genes in Japanese patients with BRCA1/2-wildtype hereditary breast and ovarian cancer syndrome. *NPJ Breast Cancer*. 2020;6:25.
11. Hiatt RA, Brody JG. Environmental determinants of breast cancer. *Annu Rev Public Health*. 2018;39:113-133.
12. Fachal L, Aschard H, Beesley J, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat Genet*. 2020;52:56-73.
13. Li H, Terry MB, Antoniou AC, et al. Alcohol consumption, cigarette smoking, and risk of breast cancer for BRCA1 and BRCA2 mutation carriers: results from the BRCA1 and BRCA2 cohort consortium. *Cancer Epidemiol Biomarkers Prev*. 2020;29:368-378.
14. Bissonauth V, Shatenstein B, Fafard E, et al. Weight history, smoking, physical activity and breast cancer risk among French-Canadian women non-carriers of more frequent BRCA1/2 mutations. *J Cancer Epidemiol*. 2009;2009:748367.
15. Hamajima N, Matsuo K, Saito T, et al. Gene-environment interactions and polymorphism studies of cancer risk in the hospital-based epidemiologic research program at Aichi Cancer Center II (HERPACC-II). *Asian Pac J Cancer Prev*. 2001;2:99-107.
16. Sueta A, Ito H, Kawase T, et al. A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast Cancer Res Treat*. 2012;132:711-721.
17. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120.
18. Zhang X, Shao Y, Tian J, et al. pTrimmer: an efficient tool to trim primers of multiplex deep sequencing data. *BMC Bioinformatics*. 2019;20:236.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
20. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491-498.
21. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
22. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44:D862-D868.
23. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56-65.
24. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285-291.
25. Nagasaki M, Yasuda J, Katsuoka F, et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. *Nat Commun*. 2015;6:8018.
26. Higasa K, Miyake N, Yoshimura J, et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. *J Hum Genet*. 2016;61:547-553.
27. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med*. 2015;34:3133-3143.
28. Mansournia MA, Geroldinger A, Greenland S, Heinze G. Separation in logistic regression: causes, consequences, and control. *Am J Epidemiol*. 2018;187:864-870.
29. Gosho M, Ohigashi T, Nagashima K, Ito Y, Maruo K. Bias in odds ratios from logistic regression methods with sparse data sets. *J Epidemiol*. 2021. doi:10.2188/jea.JE20210089. Online head of print.
30. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol*. 2016;45:1887-1894.
31. Discacciati A, Orsini N, Greenland S. Approximate Bayesian logistic regression via penalized likelihood by data augmentation. *Stata J*. 2015;15:712-736.
32. Sun J, Meng H, Yao LU, et al. Germline mutations in cancer susceptibility genes in a large series of unselected breast cancer patients. *Clin Cancer Res*. 2017;23:6113-6119.
33. Lang G-T, Shi J-X, Hu X, et al. The spectrum of BRCA mutations and characteristics of BRCA-associated breast cancers in China: Screening of 2,991 patients and 1,043 controls by next-generation sequencing. *Int J Cancer*. 2017;141:129-142.
34. Li A, Xie R, Zhi Q, et al. BRCA germline mutations in an unselected nationwide cohort of Chinese patients with ovarian cancer and healthy controls. *Gynecol Oncol*. 2018;151:145-152.
35. Wang J, Li W, Shi Y, et al. Germline mutation landscape of Chinese patients with familial breast/ovarian cancer in a panel of 22 susceptibility genes. *Cancer Med*. 2019;8:2074-2084.
36. Gao X, Nan X, Liu Y, et al. Comprehensive profiling of BRCA1 and BRCA2 variants in breast and ovarian cancer in Chinese patients. *Hum Mutat*. 2020;41:696-708.
37. Zheng W, Zhang B, Cai Q, et al. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum Mol Genet*. 2013;22:2539-2550.
38. Demir S, Tozkir H, Gurkan H, et al. Genetic screening results of individuals with high risk BRCA-related breast/ovarian cancer in Trakya region of Turkey. *J Buon*. 2020;25:1337-1347.
39. Siraj AK, Bu R, Iqbal K, et al. Prevalence, spectrum, and founder effect of BRCA1 and BRCA2 mutations in epithelial ovarian cancer from the Middle East. *Hum Mutat*. 2019;40:729-733.
40. Ahmad J, Le Calvez-Kelm F, Daud S, et al. Detection of BRCA1/2 mutations in breast cancer patients from Thailand and Pakistan. *Clin Genet*. 2012;82:594-598.
41. Inagaki-Kawata Y, Yoshida K, Kawaguchi-Sakita N, et al. Genetic and clinical landscape of breast cancers with germline BRCA1/2 variants. *Commun Biol*. 2020;3:578.
42. Drosos Y, Escobar D, Chiang M-Y, et al. ATM-deficiency increases genomic instability and metastatic potential in a mouse model of pancreatic cancer. *Sci Rep*. 2017;7:11144.
43. Timms KM, Abkevich V, Hughes E, et al. Association of BRCA1/2 defects with genomic scores predictive of DNA damage repair deficiency among breast cancer subtypes. *Breast Cancer Res*. 2014;16:475.
44. Haricharan S, Punturi N, Singh P, et al. Loss of MutL disrupts CHK2-dependent cell-cycle control through CDK4/6 to promote intrinsic endocrine therapy resistance in primary breast cancer. *Cancer Discov*. 2017;7:1168-1183.
45. Al-Ahmadie HA, Iyer G, Lee BH, et al. Frequent somatic CDH1 loss-of-function mutations in plasmacytoid variant bladder cancer. *Nat Genet*. 2016;48:356-358.
46. Boonen RACM, Rodrigue A, Stoepker C, et al. Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. *Nat Commun*. 2019;10:5296.
47. Lazaridis G, Kotoula V, Vrettou E, et al. Opposite prognostic impact of single PTEN-loss and PIK3CA mutations in early high-risk breast cancer. *Cancer Genomics Proteomics*. 2019;16:195-206.
48. Su GH, Hruban RH, Bansal RK, et al. Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers. *Am J Pathol*. 1999;154:1835-1840.
49. Harvey M, Vogel H, Morris D, Bradley A, Bernstein A, Donehower LA. A mutant p53 transgene accelerates tumour development in heterozygous but not nullizygous p53-deficient mice. *Nat Genet*. 1995;9:305-311.

50. de Vries A, Flores ER, Miranda B, et al. Targeted point mutations of p53 lead to dominant-negative inhibition of wild-type p53 function. *Proc Natl Acad Sci U S A*. 2002;99:2948-2953.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kasugai Y, Kohmoto T, Taniyama Y, et al. Association between germline pathogenic variants and breast cancer risk in Japanese women: The HERPACC study. *Cancer Sci*. 2022;113:1451–1462. doi:10.1111/cas.15312