Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Hu C, Hart SN, Gnanaolivu R, et al. A population-based study of genes previously implicated in breast cancer. N Engl J Med 2021;384:440-51. DOI: 10.1056/NEJMoa2005936

SUPPLEMENTARY APPENDIX

Contents

Acknowledgements

WHI investigators

Program Office: (National Heart, Lung, and Blood Institute, Bethesda, Maryland) Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller

Clinical Coordinating Center: (Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg

Investigators and Academic Centers: (Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Jennifer Robinson; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker; (University of Nevada, Reno, NV) Robert Brunner

Women's Health Initiative Memory Study: (Wake Forest University School of Medicine, Winston-Salem, NC) Mark Espeland

NHS/NHSII

The authors would like to thank the participants and staff of the NHS and NHSII for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

Description of Study Populations

Black Women: Etiology and Survival of Triple-negative Breast Cancers (BEST) Study: The BEST Study is a case-only study of early-onset breast cancer in Black women.[1] Black women diagnosed with invasive breast cancer at age ≤50 years between January 1, 2009 and December 31, 2012 and living in Florida at the time of their diagnosis were invited to enroll in the study. Cases were identified through the Florida State Cancer Registry. Saliva samples as a source of genomic DNA were collected at the time of enrollment.

The Black Women's Health Study (BWHS): The BWHS is an ongoing prospective follow-up study of health and illness among US black women, with a focus on cancer.[2] The study began in 1995 when 59,000 AA women 21–69 years of age from across the US completed a 14-page postal health questionnaire. Participants complete follow-up questionnaires every two years. Cases (affected) are women with incident invasive breast cancer. Controls (unaffected) are selected randomly from those who had not developed breast cancer and frequency-matched to cases on age at enrollment, most recent questionnaire completed, and geographic region of residence. Blood and saliva samples as a source of genomic DNA were obtained from 60% of all women enrolled in BWHS. Samples were obtained prior to breast cancer diagnosis for 70% of breast cancer cases with a median time to diagnosis of 6 years. The other 30% provided DNA after the breast cancer diagnosis, with a median time from diagnosis to sample collection of 5 years.

Cancer Prevention Study (CPS)-II Nutrition Cohort: The CPS-II Nutrition Cohort is a prospective cohort study of 184,194 participants (97,786 women) recruited in 1992-1993 from 21 states.[3] Most participants were aged 50-74 years and self-reported white; 1% of participants were self-reported black. At the time of recruitment, participants completed a baseline demographic, lifestyle, and medical history questionnaire. Follow-up questionnaires were sent to cohort members every two years starting in 1997 to update exposure information and to ascertain newly diagnosed cancer outcomes. Self-reported cancers were verified through abstraction of medical records or linkage with state cancer registries. DNA for this study was extracted from blood specimens donated by 21,963 female participants in 1998-2001 or from buccal cell samples from an additional 38,180 female participants in 2001-2002 for a total of 61.19% with available DNA. Among breast cancer cases, 1,871 were diagnosed prior to specimen collection with a median time between diagnosis and specimen of 3.89 years (range 0.01 – 9.36 years). A further 2,194 cases were diagnosed after specimen collection. The median time between specimen collection and diagnosis was 4.61 years (range 0.002 to 12.68 years).

Cancer Prevention Study (CPS)-3 Cohort: The CPS-3 prospective cohort includes 303,682 participants (198,632 women) who were cancer-free, aged 30-65 years and lived in 35 states or Puerto Rico at the time of recruitment which occurred between 2006 and 2013.[4] Enrollment was conducted primarily at American Cancer Society community events or at community enrollment "drives". Most women self-described as non-Hispanic white (85.3%), while 4.1% were black, 6.5% Latina, and 4.1% other. The cohort receives follow-up surveys every three years starting in 2015. For this study, we selected all self-reported breast cancer cases on the 2015 survey and a randomly-selected subcohort of the larger cohort. Nearly all women in CPS-3 (96.9%) provided a blood specimen at the time of recruitment; these specimens were the source of DNA in this study. Among cases three had diagnosis dates on or before their enrollment date. A further 2,303 cases were diagnosed after specimen collection. The median time between specimen collection and diagnosis was 2.08 years (range 0.003 to 9.31 years).

California Teachers' Study (CTS): The CTS is a prospective cohort study of 133,477 women who were residents of California and involved in the teaching professions at the time of

enrollment in 1995.[5] Follow-up surveys have been conducted on average every five years and cancer cases are identified through the California Cancer Registry. Biospecimens have been collected from 32,500 participants after enrollment. Blood and saliva samples were obtained from 107 cases prior to diagnosis with a median time between specimen collection and diagnosis of breast cancer of 37 months. Samples were collected from 2119 cases after diagnosis with a median time between diagnosis of breast cancer and specimen collection of 45 months.

Multiethnic Cohort Study (MEC): The MEC is a prospective cohort study, which enrolled 215,000 participants including 118,441 women from Los Angeles and Hawaii from 1993-1996.[6] Participants were 45-75 years of age at baseline. Follow-up data collection occurs through participant questionnaires and searches of the Hawaii and California Cancer Registry. Blood samples as a source of DNA were obtained during follow-up; samples were obtained prior to diagnosis for approximately 50% of the breast cancer cases.

Mayo Clinic Breast Cancer Study (MCBCS): The MCBCS is an on-going clinic-based case-control study initiated in February 2001 at Mayo Clinic, Rochester, MN. Cases were women over age 20 years with histologically confirmed primary invasive breast carcinoma enrolled within six months of the date of diagnosis. Controls without prior history of cancer (other than non-melanoma skin cancer) were matched on age (± 5 years) and region of residence to cases. Controls were selected from the outpatient clinic in the Department of Internal Medicine at Mayo Clinic where they were seen for general medical examinations. A self-administered risk factor questionnaire, blood sample and written informed consent were obtained from all participants at time of enrollment.[7] The proportion of women diagnosed between the ages of 50 and 75 years, the racial composition, and the distribution of clinical tumor subtype defined by hormone receptor status in MCBCS were similar to patients with breast cancer in the SEER Iowa Registry.[8]

Mayo Mammography Health Study (MMHS): The MMHS prospectively enrolled patients scheduled for a screening mammogram from October 2003 through September 2006 at the Mayo Clinic in Rochester, MN. Women were invited to take part if they were at least 35 years old, residents of Minnesota, Iowa, or Wisconsin (tri-state), and had no personal history of breast cancer. Eligible women were mailed an invitation packet consisting of a study brochure, a consent form, a baseline questionnaire, and a permission request form to link to state tumor registries. Out of 49,032 women initially invited, 10,149 were excluded for residence outside of the tri-state area (1,698), mammogram not for screening purposes (that is, a diagnostic mammogram) (6,383), and a personal history of breast cancer (2,068). Of 38,883 eligible women, 19,924 provided written informed consented (51.2% adjusted response rate) and 87% of these provided blood or saliva samples for DNA.[9] The median time from biospecimen collection to diagnosis was 6.4 years.

The Northern California Breast Cancer Family Registry (NC-BCFR): The NC-BCFR is a population-based family study conducted in Northern California, and is one of six sites collaborating in the Breast Cancer Family Registry (BCFR).[10] Cases with characteristics suggestive of inherited breast cancer were preferentially invited to enroll in the family registry; criteria included breast cancer diagnosis before age 35 years, prior ovarian or childhood cancer, bilateral breast cancer with a first diagnosis before age 50 years, or a first-degree family history of breast, ovarian, or childhood cancer. Cases that did not meet any of these criteria and were diagnosed at age 35-64 years were randomly sampled and invited to enroll. Epidemiologic data and blood or mouthwash samples were collected at enrollment. The median time from diagnosis to biospecimen collection was 26 months.

The Nurses' Health Study (NHS): The NHS is a prospective cohort study established in 1976. Married registered nurses, aged 30 to 55 in 1976, who lived in the 11 most populous states, for whom nursing boards agreed to supply NHS with member's names and addresses, were eligible to be enrolled in the cohort if they responded to the NHS baseline questionnaire. Overall, 121,700 women returned a completed questionnaire for a response rate of approximately 71% (121,700 of 172,413). Blood samples were collected from nearly 33,000 participants in 1989-90, followed by a second blood and urine sample from more than 18,700 of the same participants in 2000-02. DNA was collected from cheek cells from an additional 33,000 women in 2001-2004.[11] The median time from DNA collection to diagnosis was 10.3 years.

The Nurses' Health Study II (NHS II): The NHS II was established in 1989 to study oral contraceptives, diet, and lifestyle risk factors in a population younger than the original NHS cohort. From the 517,000 baseline questionnaires mailed, the overall response rate was approximately 24% (123,000 of 517,000). After exclusions for incomplete forms and ineligibility, a total of 116,430 women remained in NHS II cohort. Blood and urine samples from approximately 30,000 nurses, many of whom were premenopausal, were collected in 1996-1999. Over 18,500 of these women gave two blood samples and one urine sample timed within the menstrual cycle. A second blood and urine sample from approximately 16,500 of the same participants was obtained in 2010-2012, when most of the women were postmenopausal. DNA from cheek cells was collected in 2006 from another 30,000 women.[12] The median time from DNA collection to diagnosis was 7.3 years.

The Sister Study (SISTER): The Sister Study is a prospective cohort study of women who have a sister diagnosed with breast cancer.[13] The Sister Study began in July 2003, enrolling volunteers without a history of breast cancer themselves, aged 35-74, residing in the United States and Puerto Rico who had a sister with breast cancer. Enrollment of the cohort of 50,884 closed in March 2009. Baseline data collection included a comprehensive risk factor oriented Computer Assister Telephone Interview; self-administered questionnaires on family medical history, early life exposures, diet, and personal care products; and a home visit for collecting bio-specimens (first morning void urine, blood, toenail clippings, and household dust). DNA has been extracted from biospecimens for 13,432 (24.4%) of participants including 86.1% of incident breast cancer cases.

Two Sister Study (TWO SISTER): The Two Sister Study is an extension of the Sister Study that focuses on young-onset breast cancer and is based on families. Women who were younger than 50 when diagnosed with breast cancer, and had a sister enrolled in the Sister Study, were invited to participate in the Two Sister Study, along with their parents. Over 1,400 young-onset sisters enrolled in the study by completing questionnaires and/or providing saliva samples for DNA, along with 1,700 of their sisters in the Sister Study. Of their parents, 1,438 provided a saliva sample. About 1,300 of the sisters with young-onset breast cancer completed all of the study requirements (all questionnaires and saliva sample) and are now being followed prospectively along with Sister Study participants who developed breast cancer after joining the study.[14]

UCI Breast Cancer Study (UCIBCS). The UCIBCS is a case-control study in which all cases were diagnosed in Orange County, California in a 1-year period beginning 1 March 1994.[15] Eligible probands were ascertained through the population-based cancer registry of the Cancer Surveillance Program of Orange County (CSPOC) within 6 months of diagnosis. Participants signed a consent form, provided an 18-ml blood sample, and completed an epidemiological risk factor questionnaire. Controls were age-matched within 5 years of the age at diagnosis of cases.

Women's Circle of Health Study (WCHS): The WCHS is a case-control study of breast cancer initiated in four New York City boroughs and later expanded to ten counties of New Jersey.[16] Eligible cases were women with incident breast cancer diagnosed between 25 and 74 years of age; controls were identified through random-digit dialing and community health events. Blood or saliva samples as a source of genomic DNA were obtained at the time of enrollment from 94% of eligible women. The median time from diagnosis of breast cancer to blood draw was 284 days.

The Women's Health Initiative (WHI): The WHI is a prospective cohort study. Recruitment for the WHI began in 1993 and ended in 1998 and was conducted by 40 Clinical Centers in 24 states and the District of Columbia. Enrollment of racial/ethnic minority groups proportionate to the total minority population of women between 50 and 79 years of age was a high priority of the WHI. At the end of the recruitment period, 161,808 women had joined the WHI, with about 17% representing racial/ethnic minority groups.[17] Of these 87% provided biospecimens. Among 5166 women enrolled prior to diagnosis the median time from DNA collection to onset of breast cancer was 4.7 years. Among 646 women providing DNA samples after breast cancer diagnosis the median time from diagnosis to specimen collection was 1.3 years.

Wisconsin Women's Health Study (WWHS): The WWHS is a population-based case-control study conducted in Wisconsin.[18] Data for this project were collected in 2001-2007; breast cancer cases aged 20-69 years were identified from the Wisconsin mandatory cancer registry and controls were randomly selected from lists of licensed drivers and frequency-matched by age group to cases. Participants were interviewed by telephone about one year after diagnosis for cases, or a similar reference date for controls. At the conclusion of the interview, women were invited to contribute self-collected saliva samples by mail for genotyping. A total of 67.9% of eligible women provided biospecimens. The median time from breast cancer diagnosis to sample collection was 16.5 months.

Control selection for the CARRIERS study

The nested case-control studies described above (BWHS, CPSII, CTS, MEC, NHS, NSHII, and WHI), were matched on age, and had similar numbers of samples from cases and controls. BWHS selected with a 1:2 ratio of cases to controls. Differences in numbers of cases and controls in final analyses resulted from variation in DNA and sequencing quality. Controls from the CPS3 and MMHS case-cohort studies were frequency matched to cases to reflect the underlying populations in the studies. MMHS and MCBCS had an overlap in breast cancer cases. Overlapping cases were included only in MCBCS (Tables 1 and S2). Other studies (BEST, UCIBCS, NC-BCFR, SISTER and TWOSISTER) were only included in secondary analyses (Tables S1 and S2, Supplementary Appendix). UCIBCS was a case-control study enriched for younger age at diagnosis cases in which controls were frequency matched to cases. SISTER was a nested case-cohort of women with a sister with breast cancer with frequency matched controls; TWOSISTER was a familial study with controls limited to unaffected sisters of cases. NC-BCFR was a familial case-control study with a limited number of age-matched controls; BEST was a case only series and did not include any controls.

Variants cleanup and classification process

Variants were called by GATK Haplotype Caller and VarDict. Variants were viewed with VCF-Miner.[19] Pathogenic variants (PVs) were evaluated by IGV analysis of Bam files. Variants selected for study met the following criteria:

- AF (allele frequency) <0.01 in cases or <0.003 in gnomAD (or not present in gnomAD)

- AAF (altered allele frequency) >0.05 or <0.95, or =1
- Read depth of both altered reads and reference reads >5

Variant classification:
All loss of function variants (nonsense, frameshift, consensus splice sites (+/-1 or 2)) in panel genes, without an entry in Clinvar, were classified as PVs. However, truncating variants in the last 55bp of the penultimate exon or the last exon that potentially avoid nonsense mediated mRNA decay and do not influence known functional domains were excluded. Similarly variants located in genes after established cutoffs for protein function (e.g. *BRCA2* p.Tyr3208X) were excluded. All missense variants and in-frame deletions (insertion) were defined as VUS, except for first codon (Methionine) missense variants, which were classified as PVs. All intronic and synonymous variants were defined as benign.

Where variants had Clinvar classifications (from major clinical laboratories including Ambry Genetics, Invitae, GeneDX, Color Genomics and SCRP, or from the ENIGMA expert panel review), ENIGMA expert panel classification were always used. Where no expert panel results were available and there was a discrepancy in the ClinVar classifications (pathogenic or benign vs VUS) then the pathogenic or benign classifications were used.

CNV (copy number variants) were evaluated individually based on log2ratios. All suspected mosaic somatic variants (allele ratio>70:30) in *NF1* and *TP53* were excluded. Variants reported with reduced penetrance (e.g. *CHEK2* c.Ile157Thr) were excluded.

Absolute Risk estimation for Breast Cancer

Absolute risk is the probability an individual with a measured set of risk factors (e.g. mutation status) and disease-free at age a will be diagnosed with the disease in the subsequent τ years. [20] Let Z be an indicator variable for a PV in gene. We can express the absolute risk as:

$$R(a,\tau,Z) = \int_a^{a+\tau} h_1(u|Z) \times \exp\left(-\int_a^u \{h_1(v|Z) + h_2(v|Z)\}\,\mathrm{d}v\right)\mathrm{d}u$$

where $h_1(a|Z)$ is the conditional disease-specific hazard of breast cancer at age a and $h_2(a|Z)$ is the competing risks hazard at age a. The competing risk we considered is all-cause mortality. The hazard for breast cancer can be parameterized as:

$$h_1(a|Z) = h_{10}(a)\exp(\beta Z)$$

where $h_{10}(a)$ is the baseline hazard and β represent the relative risks for the risk factor Z. Estimates for β can be from the case-control data and the baseline hazards estimated by the following relationship between the baseline hazard and the marginal hazard

$$h_1^*(a) = h_{10}(a)\mathrm{E}(\exp(\beta Z)) \approx \int h_{10}(a)\exp(\beta z)dF(z)$$

where $F(Z)$ denotes the distribution of the risk factors in the population. For each age, we solve for $h_{10}(a)$ by using the expected distribution of the risk factor in the population.[21,22]

For each of the genes, we estimated absolute risk of breast cancer for a woman carrying a PV in that gene. Odds ratio estimates from affected versus unaffected women were used as estimates for $\beta(a)$ using a logistic regression model with a linear and quadratic interaction between age (in years) and mutation status to allow age-specific odds ratio estimates. A likelihood ratio test was used to evaluate the addition of the interaction parameter between age-

squared and mutation status compare to the nested model with only the interaction term between age and mutation status. The model with the quadratic interaction was statistically significantly ($p<0.05$) better for ATM, BRCA1, BRCA2, and CHEK2, but not for PALB2. The age specific odds ratios from the model where used for ages 25-85 years. SEER-21 age-specific incidence rates[23] for women diagnosed with breast cancer between 2010 and 2016 were used for estimation of the baseline hazard combined with the observed carrier frequency in the controls for the population frequency. The age-specific competing hazards model, $h_2(a|Z) \equiv h_2(a)$, was estimated from CDC US mortality rates among women, subset to the same states as the SEER-21 registries. Lifetime absolute risk curves were then estimated using the absolute risk equation above and computing cumulative risk.

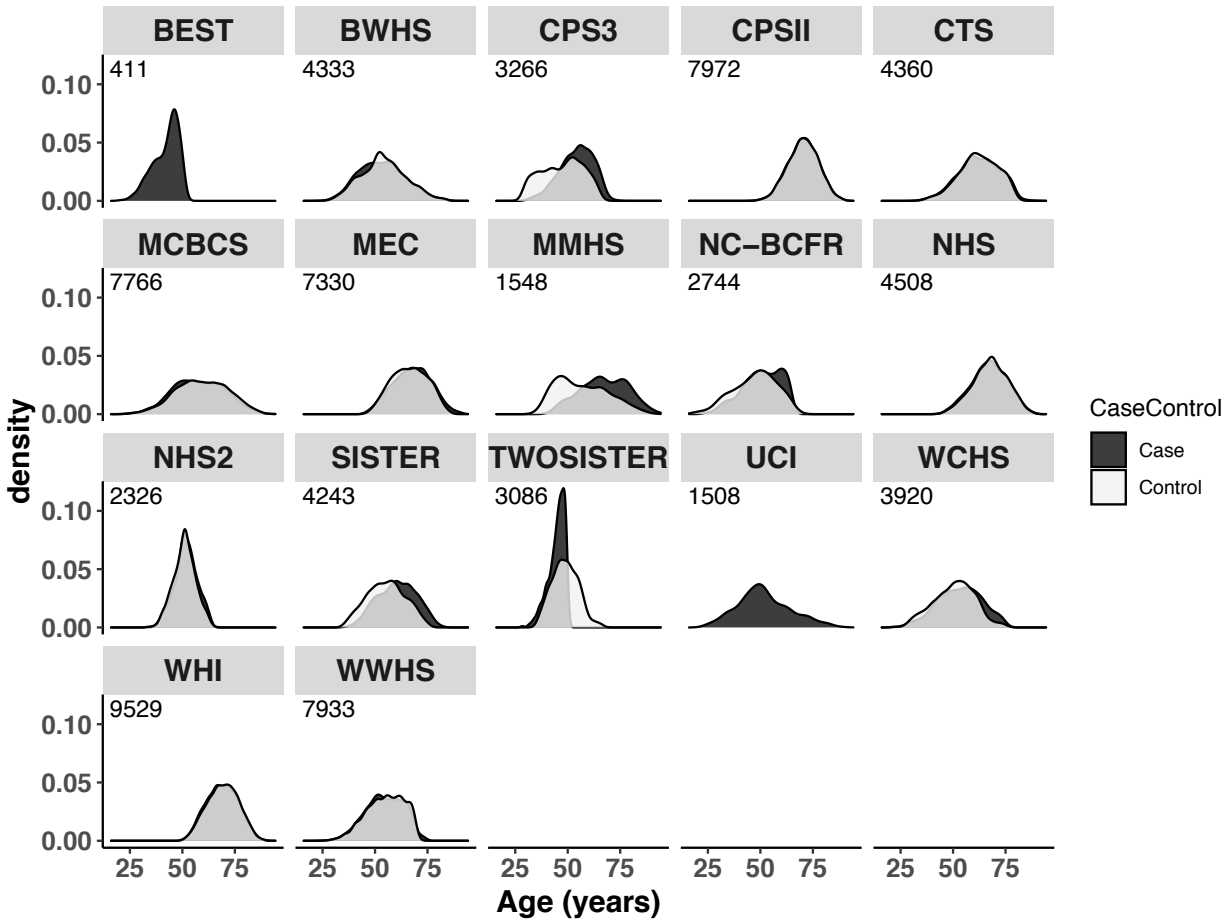

Figure S1: Age distribution of breast cancer diagnosis in cases and age at enrollment in controls from studies in the CARRIERS consortium.

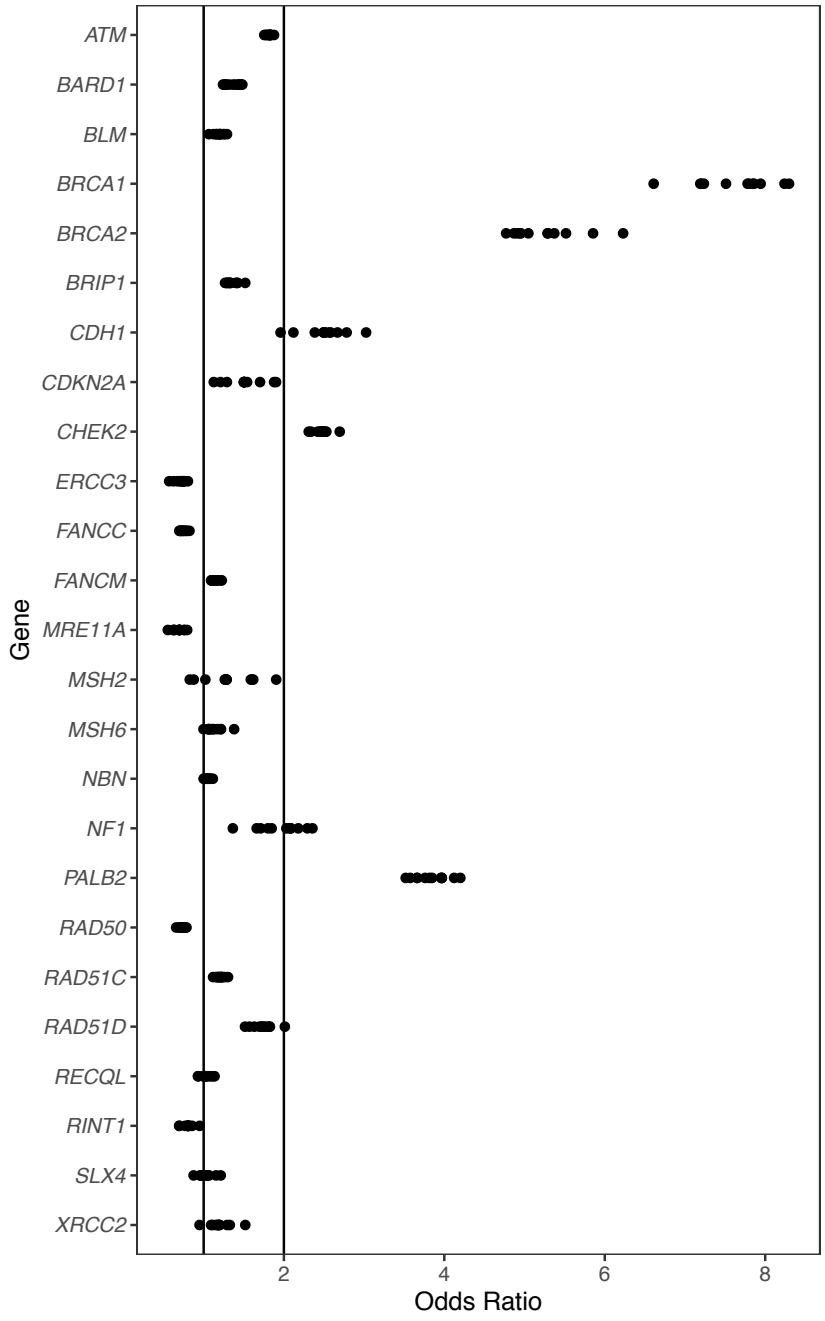

Figure S2: Sensitivity analysis for influence of each study on odds ratios for associations between PVs and breast cancer risk in the CARRIERS population-based study using a leave-one-study-out approach. Odds ratios (OR) were adjusted for study, age, family history of breast cancer, and race/ethnicity. Each point represents the gene-specific adjusted OR when a single study was removed. Results for all genes were highly consistent across studies. ORs for *ERCC3, FANCC, MRE11A,* and *RAD50* were consistently <1.0, whereas ORs for *BLM, BRIP1, CDKN2A*, and *FANCM* were consistently >1.0 suggesting that PVs in these genes may confer small, but significant increases in breast cancer risk if large enough studies were done.

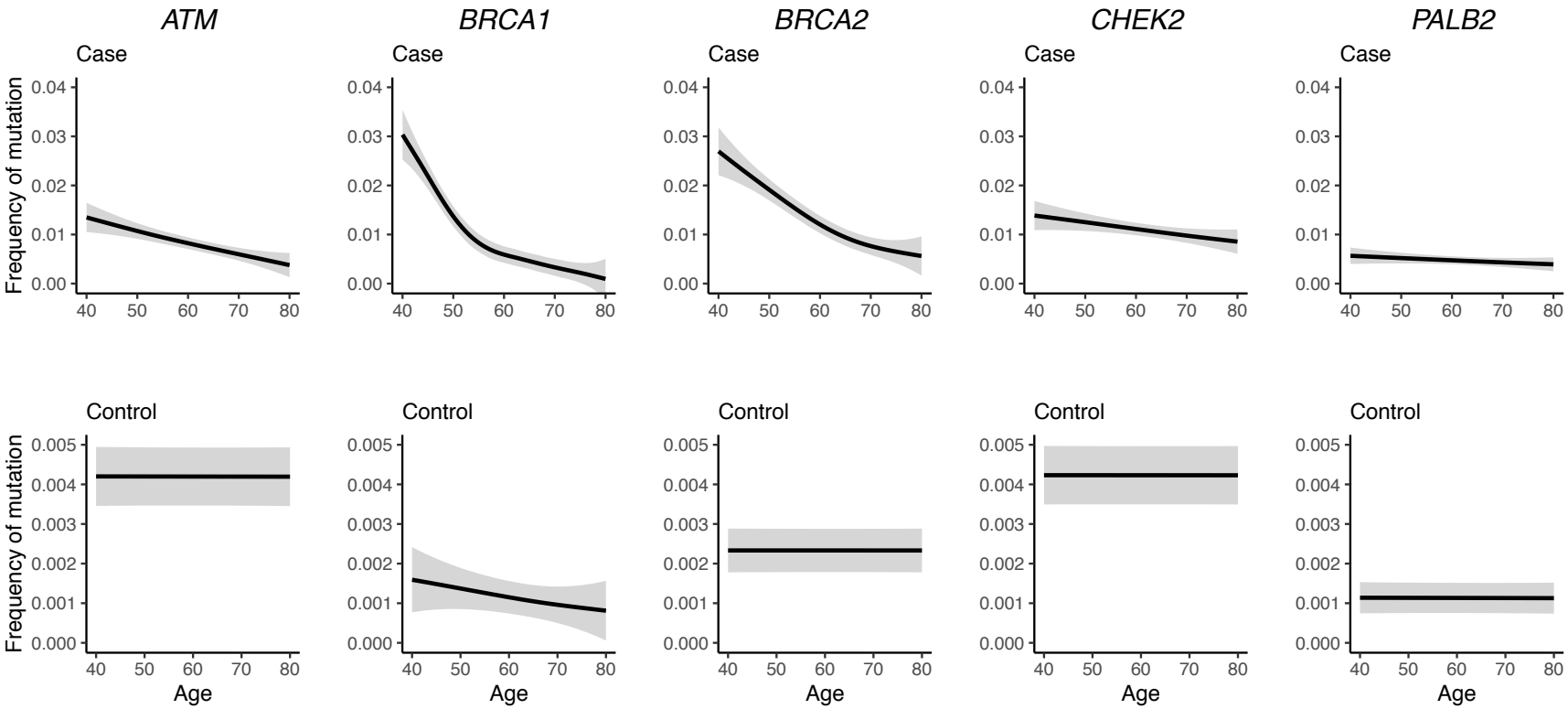

Figure S3: Frequency of PVs in the *ATM, BRCA1, BRCA2, CHEK2* and *PALB2* commonly mutated genes by age from the CARRIERS population-based study. Generalized additive model with smooth spline function for age (black line) was applied along with 95% confidence band (gray shading). Studies include BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NSHII, WCHS, WHI, and WWHS.

Table S1. List of CARRIERS consortium studies

Study Acronym	Study Name	Study design	Region of USA	Age of enrollment	% with first-degree family history of breast cancer	
					Cases	Controls
BEST	Black Women: Etiology and Survival of TNBC	Population-based case-only study	Florida	≤50	19.1	--
BWHS*	Black Women's Health Study	Prospective cohort study: nested case-control	Nationwide	21-69	15.0	9.5
CPSII*	Cancer Prevention Study-II Nutrition Cohort	Prospective cohort study: nested case-control	Nationwide	50-74	19.2	14.3
CPS3*	Cancer Prevention Study-3	Prospective cohort study: nested case-cohort	Nationwide	30-65	25.3	12.6
CTS*	California Teachers Study	Prospective cohort study: nested case-control	California	22-104	16.6	12.7
MEC*	Multiethnic Cohort	Prospective cohort study: nested case-control	California	45-75	16.5	11.2
MCBCS*	Mayo Clinic Breast Cancer Study	Hospital-based case-control study	Nationwide	≥20	24.0	20.5
MMHS*	Mayo Mammography Health Study	Prospective cohort study: nested case-cohort	MN, IA, WI	≥35	27.2	17.0
NC-BCFR	Northern California Breast Cancer Family Registry	Population-based familial case-control study	California	18-64	28.3	9.3
NHS*	Nurses' Health Study	Prospective cohort study: nested case-control	Nationwide	30-55	22.8	15.2
NHSII*	Nurses' Health Study II	Prospective cohort study: nested case-control	Nationwide	25-42	22.2	13.0
SISTER	The Sister Study	Prospective cohort study: nested case-cohort of women with a sister with breast cancer	US, Puerto Rico	35-74	100.0	100.0
TWO SISTER	Two Sister Study	Extension of The Sister Study to young-onset breast cancer from families	US, Puerto Rico	≤50	24.4	100.0
UCIBCS	UCI Breast Cancer Study	Population-based case-control study	California	40-74	--	--
WCHS*	Women's Circle of Health Study	Population-based case control study	NJ, NY	25-74	18.7	14.1
WHI*	Women's Health Initiative	Prospective cohort study: nested case-control	Nationwide	50-79	20.6	16.0
WWHS*	Wisconsin Women's Health Study	Population-based case-control study	Wisconsin	20-69	21.7	15.0

*: Studies included in CARRIERS population-based study.

Table S2: Characteristics of CARRIERS consortium participants

	Case (N=39553)	Control (N=35867)	Total (N=75420)
Age*			
N-Missing	1303	754	2057
Mean (SD)	60.44 (11.85)	60.66 (11.81)	60.55 (11.83)
Range	16.00 - 94.00	19.00 - 94.30	16.00 - 94.30
≤40	1801 (4.7%)	1771 (5.0%)	3572 (4.9%)
41-50	6670 (17.4%)	5227 (14.9%)	11897 (16.2%)
51-60	9699 (25.4%)	9572 (27.3%)	19271 (26.3%)
61-70	11631 (30.4%)	10618 (30.2%)	22249 (30.3%)
71-100	8449 (22.1%)	7925 (22.6%)	16374 (22.3%)
Race/Ethnicity			
N-Missing	190	66	256
Asian	1936 (4.9%)	1387 (3.9%)	3323 (4.4%)
African American	5287 (13.4%)	5222 (14.6%)	10509 (14.0%)
Hispanic	2208 (5.6%)	1450 (4.1%)	3658 (4.9%)
Non-Hispanic White	29250 (74.3%)	27151 (75.8%)	56401 (75.0%)
Other	682 (1.7%)	591 (1.7%)	1273 (1.7%)
Study			
BEST	411 (1.0%)	0 (0.0%)	411 (0.5%)
BWHS	1437 (3.6%)	2896 (8.1%)	4333 (5.7%)
CPSII	4037 (10.2%)	3935 (11.0%)	7972 (10.6%)
CPS3	1537 (3.9%)	1729 (4.8%)	3266 (4.3%)
CTS	2226 (5.6%)	2134 (5.9%)	4360 (5.8%)
MCBCS	4517 (11.4%)	3249 (9.1%)	7766 (10.3%)
MEC	3641 (9.2%)	3689 (10.3%)	7330 (9.7%)
MMHS	291 (0.7%)	1257 (3.5%)	1548 (2.1%)
NC-BCFR	2561 (6.5%)	183 (0.5%)	2744 (3.6%)
NHS	2088 (5.3%)	2420 (6.7%)	4508 (6.0%)
NHSII	935 (2.4%)	1391 (3.9%)	2326 (3.1%)
SISTER	2357 (6.0%)	1886 (5.3%)	4243 (5.6%)
TWO SISTER	1221 (3.1%)	502 (1.4%)	1723 (2.3%)
UCIBCS	756 (1.9%)	752 (2.1%)	1508 (2.0%)
WCHS	2215 (5.6%)	1705 (4.8%)	3920 (5.2%)
WHI	4994 (12.6%)	4535 (12.6%)	9529 (12.6%)
WWHS	4329 (10.9%)	3604 (10.0%)	7933 (10.5%)
ER status			
N-Missing	10817	NA	NA
negative	5360 (18.7%)	NA	NA
positive	23376 (81.3%)	NA	NA
PR status			
N-Missing	11564	NA	NA

negative	8335 (29.8%)	NA	NA
positive	19654 (70.2%)	NA	NA
HER2 status			
N-Missing	21830	NA	NA
negative	14632 (82.6%)	NA	NA
positive	3091 (17.4%)	NA	NA
TNBC			
N-Missing	22244	NA	NA
no	15128 (87.4%)	NA	NA
yes	2181 (12.6%)	NA	NA
Family History of BC†			
N-Missing	1811	1769	3580
no	27931 (74.0%)	27182 (79.7%)	55113 (76.7%)
yes	9811 (26.0%)	6916 (20.3%)	16727 (23.3%)
Family History of OC†			
N-Missing	5027	4469	9496
no	33242 (96.3%)	30357 (96.7%)	63599 (96.5%)
yes	1284 (3.7%)	1041 (3.3%)	2325 (3.5%)

ER: estrogen receptor. PR: progesterone receptor. HER2: human epidermal growth factor 2-neu. TNBC: triple-negative breast cancer. BC: breast cancer. OC: ovarian cancer.

*: age at breast cancer diagnosis for cases; age at selection for controls.

†: family history restricted to first-degree relatives.

Included studies: BEST, BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NC-BCFR, NHS, NHSII, SISTER, UCIBCS, WCHS, WHI, and WWHS.

Table S3: QIAseq panel genes evaluated in the CARRIERS population-based study

HGNC symbol	RefSeq	Ensemble Transcript Id	Chrom	Start position	End position	strand	Ensemble gene id
ATM	NM000051.3	ENST00000278616	11	108093211	108239829	1	ENSG00000149311
BARD1	NM000465.3	ENST00000260947	2	215590370	215674428	-1	ENSG00000138376
BLM	NM_000057.3	ENST00000355112	15	91260558	91358859	1	ENSG00000197299
BRCA1	NM007294.3	ENST00000357654	17	41196312	41277500	-1	ENSG00000012048
BRCA2	NM000059.3	ENST00000544455	13	32889611	32973805	1	ENSG00000139618
BRIP1	NM032043.2	ENST00000259008	17	59758627	59940882	-1	ENSG00000136492
CDH1	NM004360.4	ENST00000261769	16	68771128	68869451	1	ENSG00000039068
CDKN2A	NM000077.4	ENST00000304494	9	21967751	21995300	-1	ENSG00000147889
CHEK2	NM007194.3	ENST00000328354	22	29083731	29138410	-1	ENSG00000183765
ERCC3	NM000122.1	ENST00000285398	2	128014866	128051752	-1	ENSG00000163161
FANCC	NM000136.2	ENST00000289081	9	97861336	98079991	-1	ENSG00000158169
FANCM	NM020937.3	ENST00000267430	14	45605143	45670093	1	ENSG00000187790
MLH1	NM000249.3	ENST00000231790	3	37034823	37107380	1	ENSG00000076242
MRE11A	NM005591.3	ENST00000323929	11	94152895	94227074	-1	ENSG00000020922
MSH2	NM000251.2	ENST00000233146	2	47630108	47789450	1	ENSG00000095002
MSH6	NM000179.2	ENST00000234420	2	47922669	48037240	1	ENSG00000116062
NBN	NM002485.4	ENST00000265433	8	90945564	91015456	-1	ENSG00000104320
NF1	NM001042492.2	ENST00000358273	17	29421945	29709134	1	ENSG00000196712
PALB2	NM024675.3	ENST00000261584	16	23614488	23652631	-1	ENSG00000083093
PTEN	NM000314.6	ENST00000371953	10	89622870	89731687	1	ENSG00000171862
RAD50	NM_005732.3	ENST00000378823	5	131891711	131980313	1	ENSG00000113522
RAD51C	NM058216.2	ENST00000337432	17	56769934	56811703	1	ENSG00000108384
RAD51D	NM001142571	ENST00000345365	17	33426811	33448541	-1	ENSG00000185379
RECQL	NM002907	ENST00000444129	12	21621845	21654603	-1	ENSG00000004700
RINT1	NM_021930.4	ENST00000257700	7	105172532	105208124	1	ENSG00000135249
SLX4	NM_032444.2	ENST00000294008	16	3631182	3661599	-1	ENSG00000188827
TP53	NM000546.5	ENST00000269305	17	7565097	7590856	-1	ENSG00000141510
XRCC2	NM_005431.1	ENST00000359321	7	152341864	152373250	-1	ENSG00000196584

Table S4: Characteristics of CARRIERS population-based breast cancer cases and SEER-18 breast cancer cases diagnosed between 2010 and 2015

	CARRIERS population-based cases (N=32247)*	SEER18 (N=468884)
Race/Ethnicity		
N-Missing	179	9251
Asian	1282 (4.0%)	36598 (8.0%)
African American	3946 (12.3%)	51767 (11.3%)
Hispanic	1019 (3.2%)	50510 (11.0%)
Non-Hispanic-White	25287 (78.9%)	319827 (69.6%)
Other	534 (1.7%)	931 (0.2%)
Behavior		
N-Missing	1026	0
in situ	4446 (14.2%)	92594 (19.7%)
Invasive	26775 (85.8%)	376290 (80.3%)
Age of diagnosis		
N-Missing	539	17
Mean (SD)	62.07 (11.44)	61.49 (13.45)
Range	21.00 - 94.00	2.00 - 117.00
≤40	1099 (3.5%)	24770 (5.3%)
41-50	4197 (13.2%)	81573 (17.4%)
51-60	7999 (25.2%)	116418 (24.8%)
61-70	10357 (32.7%)	125208 (26.7%)
>71	8056 (25.4%)	120898 (25.8%)
ER status		
N-Missing	10014	33887
negative	3805 (17.1%)	71438 (16.4%)
positive	18428 (82.9%)	363559 (83.6%)
PR status		
N-Missing	10604	40098
negative	6186 (28.6%)	116386 (27.1%)
positive	15457 (71.4%)	312400 (72.9%)
HER2 status		
N-Missing	18995	106849
negative	11077 (83.6%)	306934 (84.8%)
positive	2175 (16.4%)	55101 (15.2%)

ER: estrogen receptor; PR: progesterone receptor; HER2: human epidermal growth factor 2-neu.
***:** Included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS.

Table S5: Associations between PVs in established and candidate predisposition genes and breast cancer risk in the overall CARRIERS study*

Gene	# of pathogenic mutation (frequency) Case (N=38332)	Control (N=35365)	OR	95% CI	P-value
Established breast cancer predisposition genes					
ATM	310 (0.81%)	146 (0.41%)	1.87	1.52-2.31	< 0.001
BARD1	56 (0.15%)	39 (0.11%)	1.36	0.89-2.11	0.16
BRCA1	442 (1.15%)	54 (0.15%)	6.43	4.78-8.84	< 0.001
BRCA2	570 (1.49%)	93 (0.26%)	5.31	4.24-6.73	< 0.001
CDH1	20 (0.05%)	6 (0.02%)	2.68	1.10-7.52	0.04
CHEK2	411 (1.07%)	160 (0.45%)	2.41	1.99-2.94	< 0.001
NF1†	27 (0.07%)	12 (0.03%)	2.16	1.05-4.74	0.04
PALB2	218 (0.57%)	43 (0.12%)	3.91	2.80-5.59	< 0.001
PTEN	9 (0.02%)	3 (0.01%)	NA	NA	NA
RAD51C	49 (0.13%)	38 (0.11%)	1.26	0.81-1.99	0.31
RAD51D	36 (0.09%)	16 (0.05%)	2.07	1.09-4.14	0.03
TP53†	24 (0.06%)	2 (0.01%)	NA	NA	NA
Total	5.67%	1.73%			
Candidate breast cancer predisposition genes					
BLM	118 (0.31%)	91 (0.26%)	1.20	0.90-1.59	0.21
BRIP1	85 (0.22%)	57 (0.16%)	1.37	0.96-1.96	0.08
CDKN2A	9 (0.02%)	5 (0.01%)	1.69	0.55-5.73	0.37
ERCC3	71 (0.19%)	98 (0.28%)	0.69	0.50-0.96	0.03
FANCC	93 (0.24%)	116 (0.33%)	0.76	0.57-1.02	0.07
FANCM	64 (0.17%)	49 (0.14%)	1.20	0.81-1.79	0.38
MLH1	12 (0.03%)	4 (0.01%)	NA	NA	NA
MRE11A	28 (0.07%)	34 (0.10%)	0.73	0.42-1.24	0.25
MSH2	9 (0.02%)	5 (0.01%)	1.46	0.47-4.95	0.52
MSH6	48 (0.13%)	36 (0.10%)	1.22	0.77-1.96	0.39
NBN	66 (0.17%)	55 (0.16%)	1.02	0.69-1.49	0.93
RAD50	72 (0.19%)	88 (0.25%)	0.76	0.54-1.06	0.10
RECQL	88 (0.23%)	78 (0.22%)	0.96	0.70-1.33	0.82
RINT1	29 (0.08%)	32 (0.09%)	0.74	0.42-1.27	0.28
SLX4	55 (0.14%)	48 (0.14%)	0.96	0.63-1.47	0.86
XRCC2	30 (0.08%)	23 (0.07%)	1.14	0.65-2.01	0.65

OR: Odds Ratio estimates for any breast cancer, adjusted for study, age, family history of breast cancer, and race/ethnicity. NA: not applicable (too few events (<5) for stable OR calculation).

*: Included studies: BEST, BWHS, CPS3, CPSII, CTS, MCBCS, MEC, MMHS, NC-BCFR, NHS, NHS2, SISTER, UCIBCS, WCHS, WHI, and WWHS.

†: *NF1, TP53* restricted to variants with AAF (altered allele frequency) between 0.3-0.7.

Table S6: Associations between PVs in established and candidate predisposition genes and breast cancer risk in CARRIERS studies with individually matched cases and controls*

Gene	# of pathogenic mutation (frequency) Case (N=17195)	Control (N=18161)	OR	95% CI	P-value
Established breast cancer predisposition genes					
ATM	122 (0.71%)	67 (0.37%)	1.99	1.44-2.74	< 0.001
BARD1	26 (0.15%)	18 (0.10%)	1.52	0.78-2.96	0.22
BRCA1	118 (0.69%)	23 (0.13%)	5.89	3.44-10.08	< 0.001
BRCA2	207 (1.20%)	46 (0.25%)	5.13	3.57-7.37	< 0.001
CDH1	9 (0.05%)	4 (0.02%)	NA	NA	NA
CHEK2	190 (1.10%)	87 (0.48%)	2.26	1.72-2.95	< 0.001
NF1†	6 (0.03%)	4 (0.02%)	NA	NA	NA
PALB2	84 (0.49%)	20 (0.11%)	3.82	2.77-5.40	< 0.001
PTEN	4 (0.02%)	0 (0.00%)	NA	NA	NA
RAD51C	19 (0.11%)	18 (0.10%)	1.08	0.55-2.11	0.82
RAD51D	12 (0.07%)	9 (0.05%)	1.48	0.62-3.52	0.38
TP53†	7 (0.04%)	1 (0.01%)	NA	NA	NA
Total	4.66%	1.64%			
Candidate breast cancer predisposition genes					
BLM	67 (0.39%)	48 (0.26%)	1.56	1.05-2.31	0.03
BRIP1	31 (0.18%)	27 (0.15%)	1.17	0.68-2.01	0.58
CDKN2A	6 (0.03%)	4 (0.02%)	1.46	0.39-5.52	0.58
ERCC3	40 (0.23%)	50 (0.28%)	0.91	0.58-1.43	0.67
FANCC	41 (0.24%)	60 (0.33%)	0.77	0.50-1.18	0.22
FANCM	30 (0.17%)	26 (0.14%)	1.46	0.83-2.58	0.19
MLH1	8 (0.05%)	0 (0.00%)	NA	NA	NA
MRE11A	14 (0.08%)	18 (0.10%)	0.60	0.26-1.42	0.24
MSH2	3 (0.02%)	2 (0.01%)	1.50	0.25-8.98	0.66
MSH6	23 (0.13%)	12 (0.07%)	2.14	1.03-4.47	0.04
NBN	28 (0.16%)	24 (0.13%)	1.21	0.68-2.15	0.52
RAD50	33 (0.19%)	50 (0.28%)	0.61	0.38-1.00	0.05
RECQL	41 (0.24%)	38 (0.21%)	1.22	0.76-1.96	0.40
RINT1	12 (0.07%)	16 (0.09%)	0.75	0.33-1.68	0.48
SLX4	23 (0.13%)	19 (0.10%)	1.36	0.72-2.57	0.35
XRCC2	4 (0.02%)	10 (0.06%)	0.31	0.09-1.14	0.08

OR:

Odds Ratio estimates for breast cancer using cases and controls matched within study by age and race/ethnicity. NA: not applicable (too few events (<5) for stable OR calculation).

*: Included studies: BWHS, CPSII, MCBCS, NHS, NHS2, TWO SISTER and WHI.

†: *NF1, TP53* restricted to variants with AAF (altered allele frequency) between 0.3-0.7.

Table S7: Predisposition gene PV frequency by race/ethnicity from the CARRIERS population-based study*

Gene	# of PV (frequency)							
	Asian (2551)		African American (8900)		Hispanic (2017)		Non-Hispanic White (50057)	
	Case (1282)	Control(1269)	Case (3946)	Control (4954)	Case (1019)	Control (998)	Case (25287)	Control (24770)
Established breast cancer predisposition gene								
ATM	4 (0.31%)	5 (0.39%)	27 (0.68%)	17 (0.34%)	9 (0.88%)	6 (0.60%)	211 (0.83%)	104 (0.42%)
BARD1	1 (0.08%)	1 (0.08%)	7 (0.18%)	8 (0.16%)	0	0	41 (0.16%)	26 (0.10%)
BRCA1	3 (0.23%)	0	41 (1.04%)	1 (0.02%)	16 (1.57%)	1 (0.10%)	212 (0.84%)	34 (0.14%)
BRCA2	8 (0.62%)	5 (0.39%)	71 (1.80%)	12 (0.24%)	14 (1.37%)	4 (0.40%)	313 (1.24%)	56 (0.23%)
CDH1	0	0	3 (0.08%)	2 (0.04%)	1 (0.10%)	1 (0.10%)	13 (0.05%)	3 (0.01%)
CHEK2	0	3 (0.24%)	15 (0.38%)	7 (0.14%)	3 (0.29%)	0	325 (1.29%)	124 (0.50%)
NF1†	0	2 (0.16%)	4 (0.10%)	1 (0.02%)	1 (0.10%)	0	14 (0.06%)	8 (0.03%)
PALB2	1 (0.08%)	1 (0.18%)	40 (1.01%)	5 (0.10%)	2 (0.20%)	1 (0.10%)	102 (0.40%)	31 (0.13%)
PTEN	0	0	1 (0.03%)	0	0	1 (0.10%)	6 (0.02%)	2 (0.01%)
RAD51C	2 (0.16%)	2 (0.16%)	7 (0.18%)	4 (0.08%)	2 (0.20%)	1 (0.10%)	29 (0.11%)	26 (0.10%)
RAD51D	1 (0.08%)	0	6 (0.15%)	3 (0.06%)	3 (0.29%)	0	15 (0.06%)	11 (0.04%)
TP53†	1 (0.08%)	0	3 (0.08%)	0	1 (0.10%)	0	13 (0.05%)	2 (0.01%)
Total	1.64%	1.60%	5.71%	1.20%	5.10%	1.50%	5.11%	1.72%
Candidate breast cancer predisposition gene								
BLM	1 (0.08%)	3 (0.24%)	6 (0.15%)	12 (0.24%)	4 (0.39%)	2 (0.20%)	93 (0.37%)	70 (0.28%)
BRIP1	1 (0.08%)	2 (0.16%)	6 (0.15%)	7 (0.14%)	2 (0.20%)	1 (0.10%)	59 (0.23%)	42 (0.17%)
CDKN2A	0	0	1 (0.03%)	0	1 (0.10%)	0	6 (0.02%)	5 (0.02%)
ERCC3	2 (0.16%)	2 (0.16%)	13 (0.33%)	9 (0.18%)	0	1 (0.10%)	40 (0.16%)	71 (0.29%)
FANCC	0	1 (0.08%)	16 (0.41%)	11 (0.22%)	0	1 (0.10%)	58 (0.23%)	90 (0.36%)
FANCM	1 (0.08%)	0	10 (0.25%)	11 (0.22%)	2 (0.20%)	1 (0.10%)	38 (0.15%)	34 (0.14%)
MLH1	1 (0.08%)	1 (0.08%)	0	1 (0.02%)	1 (0.10%)	0	8 (0.03%)	1 (0.00%)
MRE11A	0	1 (0.08%)	2 (0.05%)	2 (0.04%)	0	1 (0.10%)	23 (0.09%)	28 (0.11%)
MSH2	0	0	0	0	0	0	7 (0.03%)	5 (0.02%)
MSH6	0	1 (0.08%)	2 (0.05%)	2 (0.04%)	0	2 (0.20%)	34 (0.13%)	26 (0.10%)
NBN	0	0	4 (0.10%)	9 (0.18%)	2 (0.20%)	0	51 (0.20%)	42 (0.17%)
RAD50	2 (0.16%)	1 (0.08%)	4 (0.10%)	11 (0.22%)	5 (0.49%)	2 (0.20%)	46 (0.18%)	63 (0.25%)
RECQL	2 (0.16%)	1 (0.08%)	12 (0.30%)	5 (0.10%)	2 (0.20%)	2 (0.20%)	56 (0.22%)	59 (0.24%)
RINT1	1 (0.08%)	1 (0.08%)	2 (0.05%)	2 (0.04%)	1 (0.10%)	3 (0.30%)	18 (0.07%)	21 (0.08%)
SLX4	1 (0.08%)	3 (0.24%)	3 (0.08%)	10 (0.20%)	1 (0.10%)	0	37 (0.15%)	27 (0.11%)
XRCC2	2 (0.16%)	1 (0.08%)	3 (0.08%)	0	2 (0.20%)	1 (0.10%)	20 (0.08%)	18 (0.07%)

*: Included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS.

†: *NF1* and *TP53* restricted to PVs with AAF (altered allele frequency) between 0.3-0.7.

Table S8: Age distribution of breast cancer cases with PVs from the CARRIERS population-based study*

	Mean age (age range)				
Gene	Breast cancer cases	ER+ cases	ER- cases	With family history of breast cancer‡	Without family history of breast cancer‡
ATM	58.8 (36-86)	59.1 (36-86)	60.2 (39-77)	60.1 (38-86)	58.4 (36-83)
BARD1	59.8 (33-87)	60.7 (43-87)	61.0 (33-81)	53.2 (33-69)	62.0 (35-87)
BRCA1	50.8 (23-86)	50.9 (23-84)	50.3 (30-86)	50.9 (25-86)	50.4 (23-84)
BRCA2	55.7 (21-83)	55.4 (21-83)	58.6 (27-83)	56.6 (27-83)	55.3 (29-80)
CDH1	52.1 (27-64)	52.6 (42-63)	49.7 (27-64)	50.6 (27-63)	53.5 (42-64)
CHEK2	60.0 (28-88)	61.5 (31-88)	64.2 (44-81)	59.3 (35-88)	59.9 (28-88)
NF1†	57.8 (41-71)	58.2 (48-67)	64.4 (58-71)	52.3 (48-58)	58.8 (41-71)
PALB2	59.9 (29-84)	60.6 (29-84)	59.7 (40-80)	60.8 (39-84)	59.3 (29-82)
PTEN	52.4 (38-70)	56.7 (47-70)	NA	55.3 (48-63)	47.7 (38-53)
RAD51C	59.8 (31-83)	65.8 (45-83)	51.5 (31-71)	62.5 (51-75)	59.4 (31-83)
RAD51D	60.1 (36-82)	60.9 (39-82)	58.4 (36-74)	58.7 (39-82)	61.3 (36-82)
TP53†	48.7 (27-77)	47.9 (29-64)	64 (59-69)	44.3 (31-64)	49.9 (27-77)

NA: not applicable because of absence of PVs. ER+: estrogen receptor positive; ER-: estrogen receptor negative

*: Included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS.

†: *NF1* and *TP53* restricted to PVs with AAF (altered allele frequency) between 0.3-0.7.

‡: Family history restricted to first-degree relatives.

Table S9: VUS frequencies for 12 established breast cancer predisposition genes in the CARRIERS population-based study*

	Population-based			
	Case (N=32247)		Control (N=32544)	
Gene	# of VUS	%	# of VUS	%
ATM	1284	3.98	1245	3.83
BARD1	504	1.56	503	1.55
BRCA1	412	1.28	371	1.14
BRCA2	830	2.57	859	2.64
CDH1	488	1.51	500	1.54
CHEK2	532	1.65	499	1.53
NF1†	592	1.84	616	1.89
PALB2	532	1.65	538	1.65
PTEN	382	1.18	344	1.06
RAD51C	191	0.59	182	0.56
RAD51D	222	0.69	229	0.70
TP53†	136	0.42	143	0.44
Total	6105	18.92	6029	18.53

*: Included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NSHII, WCHS, WHI, and WWHS.

†: *NF1* and *TP53* restricted to PVs with AAF (altered allele frequency) between 0.3-0.7.

Table S10: Associations between PVs in cancer predisposition genes and breast cancer risk in the CARRIERS population-based study*

Gene	# of PV (frequency) Case (N=32247)	Control (N=32544)	OR	95% CI	P-value
Established breast cancer predisposition genes					
ATM	253 (0.78%)	134 (0.41%)	1.82	1.46-2.27	< 0.001
BARD1	49 (0.15%)	35 (0.11%)	1.37	0.87-2.16	0.18
BRCA1	275 (0.85%)	37 (0.11%)	7.62	5.33-11.27	< 0.001
BRCA2	417 (1.29%)	78 (0.24%)	5.23	4.09-6.77	< 0.001
CDH1	17 (0.05%)	6 (0.02%)	2.50	1.01-7.07	0.06
CHEK2	349 (1.08%)	138 (0.42%)	2.47	2.02-3.05	< 0.001
NF1†	19 (0.06%)	11 (0.03%)	1.93	0.91-4.31	0.09
PALB2	148 (0.46%)	38 (0.12%)	3.83	2.68-5.63	< 0.001
PTEN	8 (0.02%)	3 (0.01%)	NA	NA	NA
RAD51C	41 (0.13%)	35 (0.11%)	1.20	0.75-1.93	0.44
RAD51D	26 (0.08%)	14 (0.04%)	1.72	0.88-3.51	0.12
TP53†	19 (0.06%)	2 (0.01%)	NA	NA	NA
Total	5.01%	1.63%			
Candidate breast cancer predisposition genes					
BLM	104 (0.32%)	87 (0.27%)	1.19	0.89-1.59	0.24
BRIP1	69 (0.21%)	52 (0.16%)	1.35	0.93-1.98	0.12
CDKN2A	8 (0.02%)	5 (0.02%)	1.51	0.47-5.22	0.50
ERCC3	56 (0.17%)	83 (0.26%)	0.71	0.50-1.01	0.06
FANCC	75 (0.23%)	104 (0.32%)	0.75	0.55-1.01	0.06
FANCM	51 (0.16%)	46 (0.14%)	1.14	0.76-1.74	0.52
MLH1	10 (0.03%)	3 (0.01%)	NA	NA	NA
MRE11A	25 (0.08%)	32 (0.10%)	0.69	0.38-1.20	0.19
MSH2	7 (0.02%)	5 (0.02%)	1.28	0.38-4.47	0.68
MSH6	39 (0.12%)	32 (0.10%)	1.13	0.70-1.83	0.63
NBN	57 (0.18%)	51 (0.16%)	1.05	0.71-1.56	0.81
RAD50	57 (0.18%)	82 (0.25%)	0.73	0.51-1.04	0.08
RECQL	74 (0.23%)	69 (0.21%)	1.03	0.74-1.45	0.86
RINT1	24 (0.07%)	28 (0.09%)	0.80	0.45-1.41	0.44
SLX4	44 (0.14%)	41 (0.13%)	1.03	0.66-1.60	0.91
XRCC2	27 (0.08%)	21 (0.06%)	1.19	0.67-2.17	0.55

OR: Odds Ratio estimates for any breast cancer, adjusted for study, age, FH of breast cancer, and race/ethnicity. NA: not applicable (too few events (<5) for stable OR calculation).
*: Included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS.
†: *NF1, TP53* restricted to PVs with AAF (altered allele frequency) between 0.3-0.7.

Table S11: Associations between PVs in candidate predisposition genes and breast cancer risk by ER status of tumors from the CARRIERS population-based study*

	ER+ (N=18428)			ER- (N=3805)			TNBC (N=1463)		
Gene	# of PV (Freq)	OR (95%CI)	P-value	# of PV (Freq)	OR (95%CI)	P-value	# of PV (Freq)	OR (95%CI)	P-value
BLM	65 (0.35%)	1.33 (0.94-1.87)	0.11	9 (0.24%)	0.98 (0.46-1.87)	0.96	2 (0.14%)	NA	NA
BRIP1	36 (0.20%)	1.32 (0.83-2.08)	0.23	9 (0.24%)	1.42 (0.61-2.90)	0.38	5 (0.34%)	1.80 (0.53-4.61)	0.28
CDKN2A	2 (0.01%)	NA	NA	0 (0.00%)	NA	NA	0 (0.00%)	NA	NA
ERCC3	36 (0.20%)	0.82 (0.53-1.23)	0.34	6 (0.16%)	0.75 (0.29-1.59)	0.49	2 (0.14%)	NA	NA
FANCC	44 (0.24%)	0.80 (0.54-1.16)	0.24	10 (0.26%)	0.89 (0.43-1.65)	0.73	4 (0.27%)	NA	NA
FANCM	24 (0.13%)	1.18 (0.68-2.01)	0.55	10 (0.26%)	1.57 (0.70-3.22)	0.24	5 (0.34%)	1.71 (0.50-4.49)	0.33
MLH1	4 (0.02%)	NA	NA	3 (0.08%)	NA	NA	1 (0.07%)	NA	NA
MRE11A	13 (0.07%)	0.67 (0.32-1.33)	0.27	5 (0.13%)	0.60 (0.10-2.00)	0.48	1 (0.07%)	NA	NA
MSH2	4 (0.02%)	NA	NA	1 (0.03%)	NA	NA	1 (0.07%)	NA	NA
MSH6	22 (0.12%)	0.98 (0.54-1.77)	0.94	7 (0.18%)	1.84 (0.73-4.07)	0.16	3 (0.21%)	NA	NA
NBN	32 (0.17%)	1.21 (0.73-1.98)	0.45	4 (0.11%)	NA	NA	1 (0.07%)	NA	NA
RAD50	37 (0.20%)	0.80 (0.52-1.20)	0.29	7 (0.18%)	0.65 (0.25-1.39)	0.32	4 (0.27%)	NA	NA
RECQL	39 (0.21%)	0.96 (0.62-1.46)	0.84	11 (0.29%)	1.46 (0.72-2.70)	0.26	4 (0.27%)	NA	NA
RINT1	16 (0.09%)	0.80 (0.41-1.53)	0.57	2 (0.05%)	NA	NA	1 (0.07%)	NA	NA
SLX4	23 (0.13%)	0.83 (0.48-1.42)	0.51	6 (0.16%)	1.03 (0.39-2.31)	0.95	1 (0.07%)	NA	NA
XRCC2	8 (0.04%)	1.09 (0.42-2.70)	0.86	4 (0.11%)	NA	NA	2 (0.14%)	NA	NA

OR: Odds Ratio estimates for breast cancer were adjusted for study, age, family history of breast cancer, and race/ethnicity. NA: not applicable.

*: Included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS.

Table S12: Associations between PVs and breast cancer risk in the CARRIERS population-based study for cases with or without family history of breast cancer*

	Family history of breast cancer (N=6361)				No family history of breast cancer (N=24873)			
Gene	**Cases with PVs (Freq)**	**OR**	**95% CI**	**P value**	**Cases with PVs (Freq)**	**OR**	**95% CI**	**P value**
Established breast cancer predisposition gene								
ATM	61 (0.96%)	2.15	1.56-2.93	< 0.001	183 (0.74%)	1.72	1.37-2.16	<0.001
BARD1	9 (0.14%)	1.36	0.61-2.74	0.41	36 (0.14%)	1.38	0.86-2.21	0.18
BRCA1	113 (1.78%)	17.15	11.83-25.43	< 0.001	150 (0.60%)	5.45	3.82-7.98	<0.001
BRCA2	157 (2.47%)	10.71	8.12-14.24	< 0.001	250 (1.01%)	4.25	3.30-5.54	<0.001
CDH1	8 (0.13%)	6.56	2.24-20.35	< 0.001	9 (0.04%)	1.85	0.64-5.71	0.26
CHEK2	97 (1.52%)	3.59	2.75-4.68	< 0.001	237 (0.95%)	2.25	1.81-2.79	<0.001
NF1†	3 (0.05%)	NA	NA	NA	16 (0.06%)	2.11	0.97-4.81	0.07
PALB2	60 (0.94%)	8.04	5.33-12.29	< 0.001	86 (0.35%)	2.97	2.03-4.42	<0.001
PTEN	2 (0.03%)	NA	NA	NA	5 (0.02%)	NA	NA	NA
RAD51C	5 (0.08%)	0.71	0.24-1.67	0.48	36 (0.14%)	1.27	0.79-2.05	0.32
RAD51D	7 (0.11%)	2.49	0.93-6.06	0.05	18 (0.07%)	1.53	0.75-3.16	0.24
TP53†	5 (0.08%)	NA	NA	NA	14 (0.06%)	NA	NA	NA
Candidate breast cancer predisposition gene								
BLM	23 (0.36%)	1.38	0.85-2.15	0.18	80 (0.32%)	1.22	0.90-1.66	0.21
BRIP1	20 (0.31%)	2.15	1.25-3.58	0.004	47 (0.19%)	1.20	0.80-1.81	0.37
CDKN2A	0 (0.00%)	NA	NA	NA	7 (0.03%)	2.00	0.62-6.98	0.25
ERCC3	9 (0.14%)	0.55	0.26-1.05	0.10	46 (0.18%)	0.75	0.52-1.09	0.13
FANCC	13 (0.20%)	0.61	0.32-1.06	0.10	61 (0.25%)	0.78	0.56-1.08	0.14
FANCM	7 (0.11%)	0.75	0.31-1.56	0.48	43 (0.17%)	1.19	0.78-1.81	0.42
MLH1	3 (0.05%)	NA	NA	NA	7 (0.03%)	2.61	0.72-12.12	0.17
MRE11A	3 (0.05%)	NA	NA	NA	17 (0.07%)	0.74	0.40-1.33	0.32
MSH2	2 (0.03%)	NA	NA	NA	5 (0.02%)	1.33	0.37-4.84	0.65
MSH6	8 (0.13%)	1.26	0.54-2.62	0.57	30 (0.12%)	1.11	0.67-1.84	0.69
NBN	12 (0.19%)	1.09	0.54-2.03	0.80	41 (0.16%)	1.08	0.71-1.64	0.72
RAD50	11 (0.17%)	0.71	0.35-1.27	0.28	44 (0.18%)	0.71	0.48-1.02	0.07
RECQL	8 (0.13%)	0.47	0.20-0.96	0.06	63 (0.25%)	1.19	0.84-1.68	0.33
RINT1	2 (0.03%)	NA	NA	NA	21 (0.08%)	0.95	0.53-1.69	0.86
SLX4	9 (0.14%)	1.00	0.43-2.04	1.00	35 (0.14%)	1.07	0.67-1.69	0.77
XRCC2	5 (0.08%)	1.06	0.35-2.61	0.91	22 (0.09%)	1.22	0.67-2.23	0.52

OR: Odds Ratio estimates for breast cancer were adjusted for study, age, and race/ethnicity. NA: not applicable (too few PVs in case or control (<5) for stable OR calculation).

***:** Family history refers to first degree relatives with breast cancer. Included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS. Controls were all unaffected participants in CARRIERS population-based studies (Table 1).

†: *NF1* and *TP53* restricted to PVs with AAF (altered allele frequency) between 0.3-0.7.

Table S13: Associations between predisposition gene PVs and risk of breast cancer diagnosed at ≤ 50 years of age in the CARRIERS population-based study*

	# of PV (frequency)				
Gene	Case (N=5296)	Control (N=6042)	OR	95%CI	p-value
ATM	67 (1.27%)	28 (0.46%)	2.30	1.46-3.71	< 0.001
BARD1	9 (0.17%)	8 (0.13%)	0.94	0.34-2.61	0.90
BRCA1	144 (2.72%)	11 (0.18%)	16.14	8.50-34.96	< 0.001
BRCA2	135 (2.55%)	14 (0.23%)	11.79	6.83-22.18	< 0.001
CHEK2	74 (1.40%)	26 (0.43%)	2.64	1.67-4.31	< 0.001
PALB2	31 (0.59%)	11 (0.18%)	2.91	1.46-6.27	0.004
RAD51C	9 (0.17%)	5 (0.08%)	2.75	0.89-9.77	0.09

OR: Odds Ratio estimates for breast cancer were adjusted for study, age, and family history of breast cancer, and race/ethnicity.

*: Analyses of *CDH1, NF1, PTEN, RAD51D*, and *TP53* were not performed because of limiting numbers of PVs; included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS.

Table S14. Associations between predisposition gene PVs and risk of breast cancer diagnosed at >50 years of age in the CARRIERS population-based study*

	# of PV (frequency)				
Gene	Case (N=26412)	Control (N=26500)	OR	95%CI	p-value
ATM	181 (0.69%)	106 (0.40%)	1.68	1.31-2.17	< 0.001
BARD1	40 (0.15%)	27 (0.10%)	1.44	0.87-2.42	0.16
BRCA1	127 (0.48%)	26 (0.10%)	4.61	2.99-7.42	< 0.001
BRCA2	265 (1.00%)	64 (0.24%)	3.97	3.02-5.30	< 0.001
CDH1	11 (0.04%)	5 (0.02%)	2.09	0.74-6.78	0.18
CHEK2	271 (1.03%)	112 (0.42%)	2.35	1.88-2.97	< 0.001
NF1†	15 (0.06%)	8 (0.03%)	1.94	0.83-4.90	0.14
PALB2	115 (0.44%)	27 (0.10%)	4.23	2.78-6.70	< 0.001
RAD51C	30 (0.11%)	30 (0.11%)	1.01	0.60-1.70	0.96
RAD51D	18 (0.07%)	10 (0.04%)	1.73	0.81-3.92	0.17

OR: Odds Ratio estimates for breast cancer were adjusted for study, age, and family history of breast cancer, and race/ethnicity.

*: Analyses of *PTEN* and *TP53* were not performed because of limiting numbers of mutations; included studies: BWHS, CPSII, CPS3, CTS, MCBCS, MEC, MMHS, NHS, NHSII, WCHS, WHI, and WWHS.

†: *NF1* restricted to PVs with AAF (altered allele frequency) between 0.3-0.7.

Table S15: Associations between predisposition genes PVs and risk of breast cancer in CARRIERS population-based studies with enrollment at <45 years of age*

Gene	# of PV (frequency)		OR	95% CI	P value
	Case (N=17487)	Control (N=17965)			
Established breast cancer predisposition gene					
ATM	162 (0.93%)	80 (0.45%)	1.89	1.43-2.53	<0.001
BARD1	33 (0.19%)	22 (0.12%)	1.37	0.78-2.43	0.28
BRCA1	209 (1.20%)	27 (0.15%)	8.63	5.63-13.89	<0.001
BRCA2	296 (1.69%)	38 (0.21%)	7.65	5.47-11.02	<0.001
CDH1	15 (0.09%)	5 (0.03%)	2.66	1.00-8.38	0.066
CHEK2	218 (1.25%)	72 (0.40%)	3.06	2.32-4.08	<0.001
NF1†	13 (0.07%)	6 (0.03%)	2.68	1.01-7.95	0.06
PALB2	89 (0.51%)	22 (0.12%)	3.99	2.50-6.67	<0.001
PTEN	7 (0.04%)	3 (0.02%)	NA	NA	NA
RAD51C	26 (0.15%)	20 (0.11%)	1.26	0.69-2.35	0.45
RAD51D	16 (0.09%)	6 (0.03%)	2.41	0.91-7.60	0.097
TP53†	13 (0.07%)	0 (0.00%)	NA	NA	NA
Total					
Candidate breast cancer predisposition gene					
BLM	56 (0.32%)	53 (0.30%)	1.06	0.72-1.57	0.76
BRIP1	41 (0.23%)	35 (0.19%)	1.22	0.75-1.99	0.42
CDKN2A	4 (0.02%)	2 (0.01%)	NA	NA	NA
ERCC3	29 (0.17%)	46 (0.26%)	0.67	0.41-1.10	0.12
FANCC	40 (0.23%)	58 (0.32%)	0.74	0.48-1.12	0.16
FANCM	31 (0.18%)	30 (0.17%)	1.12	0.66-1.92	0.67
MLH1	4 (0.02%)	2 (0.01%)	NA	NA	NA
MRE11A	15 (0.09%)	16 (0.09%)	0.95	0.44-2.01	0.90
MSH2	5 (0.03%)	3 (0.02%)	NA	NA	NA
MSH6	27 (0.15%)	17 (0.09%)	1.39	0.75-2.63	0.30
NBN	30 (0.17%)	33 (0.18%)	0.88	0.52-1.49	0.64
RAD50	29 (0.17%)	44 (0.24%)	0.67	0.41-1.09	0.11
RECQL	45 (0.26%)	35 (0.19%)	1.20	0.76-1.92	0.43
RINT1	13 (0.07%)	13 (0.07%)	0.90	0.39-2.04	0.79
SLX4	27 (0.15%)	23 (0.13%)	1.10	0.62-1.98	0.74
XRCC2	21 (0.12%)	13 (0.07%)	1.37	0.69-2.83	0.38

OR: Odds Ratio estimates for breast cancer were adjusted for study, age, family history of breast cancer, and race/ethnicity; NA: not applicable (too few events (<5) for stable OR calculation).

***:** Studies with minimum age of enrollment <45 years: BWHS, CPS3, CTS, MCBCS, MMHS, NHSII, WCHS, and WWHS.

†: *NF1* and *TP53* restricted to PVs with AAF (altered allele frequency) between 0.3-0.7.

References

1. Pal T, Bonner D, Cragun D, et al. A high frequency of BRCA mutations in young black women with breast cancer residing in Florida. Cancer 2015;121:4173-80.
2. Palmer JR, Ruiz-Narvaez EA, Rotimi CN, et al. Genetic susceptibility loci for subtypes of breast cancer in an African American population. Cancer Epidemiol Biomarkers Prev 2013;22:127-34.
3. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer 2002;94:500-11.
4. Patel AV, Jacobs EJ, Dudas DM, et al. The American Cancer Society's Cancer Prevention Study 3 (CPS-3): Recruitment, study design, and baseline characteristics. Cancer 2017;123:2014-24.
5. Bernstein L, Allen M, Anton-Culver H, et al. High breast cancer incidence rates among California teachers: results from the California Teachers Study (United States). Cancer Causes Control 2002;13:625-35.
6. Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151:346-57.
7. Vachon CM, Li J, Scott CG, et al. No evidence for association of inherited variation in genes involved in mitosis and percent mammographic density. Breast Cancer Res 2012;14:R7.
8. Yadav S, Hu C, Hart SN, et al. Evaluation of Germline Genetic Testing Criteria in a Hospital-Based Series of Women With Breast Cancer. Journal of Clinical Oncology;0:JCO.19.02190.
9. Olson JE, Sellers TA, Scott CG, et al. The influence of mammogram acquisition on the mammographic density and breast cancer association in the Mayo Mammography Health Study cohort. Breast Cancer Res 2012;14:R147.
10. John EM, Sangaramoorthy M, Koo J, Whittemore AS, West DW. Enrollment and biospecimen collection in a multiethnic family cohort: the Northern California site of the Breast Cancer Family Registry. Cancer Causes Control 2019;30:395-408.
11. Eckel N, Li Y, Kuxhaus O, Stefan N, Hu FB, Schulze MB. Transition from metabolic healthy to unhealthy phenotypes and association with cardiovascular disease risk across BMI categories in 90 257 women (the Nurses' Health Study): 30 year follow-up from a prospective cohort study. Lancet Diabetes Endocrinol 2018;6:714-24.
12. Hirko KA, Chai B, Spiegelman D, et al. Erythrocyte membrane fatty acids and breast cancer risk: a prospective analysis in the nurses' health study II. Int J Cancer 2018;142:1116-29.
13. Sandler DP, Hodgson ME, Deming-Halverson SL, et al. The Sister Study Cohort: Baseline Methods and Participant Characteristics. Environ Health Perspect 2017;125:127003.
14. Fei C, Deroo LA, Sandler DP, Weinberg CR. Fertility drugs and young-onset breast cancer: results from the Two Sister Study. J Natl Cancer Inst 2012;104:1021-7.
15. Anton-Culver H, Cohen PF, Gildea ME, Ziogas A. Characteristics of BRCA1 mutations in a population-based case series of breast and ovarian cancer. Eur J Cancer 2000;36:1200-8.
16. Ambrosone CB, Ciupak GL, Bandera EV, et al. Conducting Molecular Epidemiological Research in the Age of HIPAA: A Multi-Institutional Case-Control Study of Breast Cancer in African-American and European-American Women. J Oncol 2009;2009:871250.
17. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:629-42.
18. Trentham-Dietz A, Sprague BL, Hampton JM, et al. Modification of breast cancer risk according to age and menopausal status: a combined analysis of five population-based case-control studies. Breast Cancer Res Treat 2014;145:165-75.
19. Hart SN, Duffy P, Quest DJ, Hossain A, Meiners MA, Kocher JP. VCF-Miner: GUI-based application for mining variants and annotations stored in VCF files. Brief Bioinform 2016;17:346-51.

20. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879-86.
21. Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst 1999;91:1541-8.
22. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 2016;17:392-406.
23. National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch. Surveillance, Epidemiology, and End Results (SEER) Program Research Data (1973-2014). April 2017 (Based on November 2016 Submission).