HGG Advances

ARTICLE

Associations of genetic susceptibility to 16 cancers with risk of breast cancer overall and by intrinsic subtypes

Jungyoon Choi,[1] Guochong Jia,[1] Wanqing Wen,[1] Ran Tao,[2] Jirong Long,[1] Xiao-Ou Shu,[1] and Wei Zheng[1,*]

Summary

Certain genetic variants are associated with risks of multiple cancers. We investigated breast cancer risk with overall genetic susceptibility to each of 16 other cancers. We constructed polygenic risk scores (PRS) for 16 cancers using risk variants identified by genome-wide association studies. We evaluated the associations of these PRSs with breast cancer risk (overall and by subtypes) using Breast Cancer Association Consortium data, including 106,278 cases and 91,477 controls of European ancestry. Odds ratios (OR) and 95% confidence intervals (CIs) were estimated to measure the association of each PRS with breast cancer risk. Data from the UK Biobank, including 4,337 cases and 209,983 non-cases, were used to replicate the findings. A 5%–8% significantly elevated risk of overall breast cancer was associated with per unit increase of the PRS for glioma and cancers of the corpus uteri, stomach, or colorectum. Analyses by subtype revealed that the PRS for corpus uteri cancer (OR = 1.09; 95% CI, 1.03–1.15) and stomach cancer (OR = 1.07; 95% CI, 1.03–1.12) were associated with estrogen receptor-positive breast cancer, while ovarian cancer PRS was associated with triple-negative breast cancer (OR = 1.25; 95% CI, 1.01–1.55). UK Biobank data supported the positive associations of overall breast cancer risk with PRS for melanoma and cancers of the stomach, colorectum, and ovary. Our study provides strong evidence for shared genetic susceptibility of breast cancer with several other cancers. Results from our study help uncover the genetic basis for breast and other cancers and identify individuals at high risk for multiple cancers.

Introduction

Some genome-wide association studies (GWAS)-identified risk variants are shared across multiple cancers. For example, genetic variants at chromosome 8q24 were found to be associated with cancers of prostate (MIM: 176807), colorectum (MIM: 114500), breast (MIM: 114480), bladder (MIM: 109800), and other sites;[1–7] and genetic variants in and near the telomerase reverse transcriptase (*TERT*) (MIM: 187270) gene were associated with glioma (MIM: 137800) and cancers of the lung (MIM: 211980), breast, and colorectum.[5,8–11] Several studies have estimated correlation of genetic risks across cancer types.[12–15] Sampson et al. evaluated 13 cancers in populations of European ancestry and found four cancer pairs with marginally significant correlations in genetic risk (kidney [MIM: 144700] and testes [MIM: 273300]; diffuse large B cell lymphoma [MIM: 605027] and pediatric osteosarcoma [MIM: 259500]; diffuse large B cell lymphoma and chronic lymphocytic leukemia [MIM: 151400]; bladder and lung).[12] Jiang et al. observed four statistically significant correlations in genetic risks (lung and head/neck cancer [MIM: 275355]; breast and ovarian cancer [MIM: 167000]; breast and lung cancer; breast and colorectal cancer).[13] Using polygenic risk score (PRS) as a measure of the cumulative effect of risk variants identified for a cancer, we tested the hypothesis that the overall genetic susceptibility to certain cancers may be related to breast cancer risk. Given that breast cancer is a molecularly diverse disease, we evaluated further the association by breast cancer intrinsic subtypes defined by estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), as well as tumor grade.[16] We systematically evaluated the association of PRS for each of 16 major cancers in relation to breast cancer risk, overall and by subtypes, using data obtained from more than 400,000 women of European ancestry, including more than 110,000 cases of breast cancer.

Material and methods

Data sources

We acquired summary-level statistics data generated from 197,755 women (106,278 cases of breast cancer) of European ancestry included in the Breast Cancer Association Consortium (BCAC) (BCAC Data: http://bcac.ccge.medschl.cam.ac.uk/; Table S1). The design and methods of the BCAC have been described previously.[17] We used individual-level data from 214,320 women (4,337 cases of incident breast cancer) of European ancestry from the UK Biobank (UK Biobank Data: https://www.ukbiobank.ac.uk/), a large prospective population-based cohort study, to replicate our findings based on the BCAC data. The design and methods of the UK Biobank study have been described previously.[18] In the UK Biobank, data on the

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN 37203, USA; [2]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37203, USA
*Correspondence: wei.zheng@vanderbilt.edu
https://doi.org/10.1016/j.xhgg.2021.100077.

diagnosis of cancers was provided by the National Health Service (NHS) Information Center for participants from England and Wales (follow-up through 31 March 2016) and by the NHS Central Register Scotland for participants from Scotland (follow-up through 31 October 2015). Cancer codes were from the International Classification of Diseases, Ninth Revision (ICD-9) or the International Classification of Diseases, Tenth Revision (ICD-10). Only the first diagnosed malignant tumors other than non-melanoma skin cancer (C44 in ICD-10, or 173 in ICD-9) were considered in this study. The outcome for this study is incident breast cancer as the first cancer diagnosis with codes of ICD-9 = 174 or ICD-10 = C50. The study was approved by the ethical committee at Vanderbilt University Medical Center, and all participating studies were approved by ethical committees of their institutions.

Genotyping and imputation

In the BCAC, details on genotype calling, quality control, and imputation were described previously.[17,19–22] After quality control, variants were imputed using the 1000 Genomes Project phase 3. We obtained imputed genotype data from 487,154 participants in the UK Biobank. Samples were genotyped using two arrays sharing 95% marker content: the UK BiLEVE Axiom (UKBL; 807,411 markers) and the UK Biobank Axiom (UKBB; 825,927 markers). These genotyping data were imputed using reference panels of the Haplotype Reference Consortium, or UK10K, and 1000 Genomes Project phase 3. European ancestry of study participants was determined using the genotype data by projecting all of the UK Biobank samples on the first two major principal components of four populations included in the 1000 Genomes Project (CEU, YRI, CHB, and JPT).[23] Individuals not falling in the CEU cluster were excluded (n = 23,409). Those self-reporting as non-European were also excluded (n = 4,916). In the dataset from the UK Biobank, a kinship coefficient was estimated for each pair of samples using KING's robust estimator.[24] We excluded second-degree (or higher) related individuals (kinship coefficient $\geq$ 0.0442; n = 35,067). We excluded participants who had been diagnosed with cancer before the beginning of the study—which was the baseline (n = 22,759) —and those aged below 40 years (n = 5). After these exclusions, 400,610 individuals (186,290 men and 214,320 women) remained for the current analysis. We included 214,320 women for replication analyses of the association of 16 cancer-specific PRSs with breast cancer risk.

Selection of known cancer susceptibility variants

Known susceptibility variants associated with breast cancer and 16 other cancers were selected by reviewing the GWAS catalog and PubMed publications. The 16 other cancers evaluated in this study included cancers of the bladder, colorectum, corpus uteri (MIM: 608089), esophagus (MIM: 133239), kidney, lung, ovary, pancreas (MIM: 260350), prostate and stomach (MIM: 613659), glioma, melanoma (MIM: 155600), and hematologic malignancies (chronic lymphoid leukemia; diffuse large B cell lymphoma; follicular lymphoma [MIM: 605027]; and multiple myeloma [MIM: 254500]). We selected genetic risk variants, including single-nucleotide polymorphisms (SNPs) or small insertions or deletions from the most recent studies with the largest sample sizes of individuals of European ancestry (sample size varied from 5,415 to 299,686; Table S2).[25,26] Using the conventional genome-wide significance threshold ($p < 5 \times 10^{-8}$), variants showing an association with p values at or below this threshold were included in our study. We also included some risk variants with an established association at $p < 5 \times 10^{-8}$ from previous studies with the cancer of interest even if they were not significant at $p < 5 \times 10^{-8}$ in the latest studies due to small sample sizes. Cancer risk variants on the X chromosome and those reported exclusively from non-European populations were excluded from this study. For variants in linkage disequilibrium (LD) ($r^2 \geq 0.2$) with each other in European ancestry populations in the 1000 Genomes Project, only the variant with the lowest p value was included in this study. In total, from previously reported GWAS data, we selected 497 unique risk variants for 16 cancers of interest (Table S3). We further applied more stringent criteria to select variants (MAF > 0.01, $r^2 < 0.01$ for LD), leaving 456 variants associated with 16 types of cancer to construct a PRS for each cancer. Variants were selected based on an imputation quality score >0.8 from both BCAC and UK Biobank. Final variants used from the BCAC dataset are shown in Table S4 and final variants used from the UK Biobank dataset are shown in Table S5.

Statistical analyses

We only included participants of European ancestry from the BCAC and UK Biobank in this study, as the PRSs were derived using risk variants for 16 cancers identified from GWAS conducted in this population. The overall study design is outlined in Figure 1. Outcome variables included the overall breast cancer risk, as well as its subtype according to the ER status (ER-positive and ER-negative), and five intrinsic subtypes based on the combined tumor status and grade (luminal A: ER+ and/or PR+, HER2−, grades 1 and 2; luminal B/HER2-negative: ER+ and/or PR+, HER2−, grade 3; luminal B: ER+ and/or PR+, HER2+; HER2-enriched: ER− and PR−, HER2+; triple-negative: ER−, PR−, HER2−).[17]

To study the association between each cancer PRS (trait) and risk of breast cancer overall and by its subtypes (outcome), inverse-variance weighted (IVW) meta-analyses with a random-effect model were performed on each cancer type except for gastric cancer (which was analyzed with a fixed-effect model because there were only three variants for gastric cancer).[27,28] We obtained beta coefficients and standard errors for each SNP-trait association from previous GWAS publications; we extracted the same statistics for each SNP-outcome (breast cancer risk) association from the BCAC data (Table S4). We harmonized the variants (matching effect alleles) across the two datasets counting the allele, which was associated with an increased trait risk. Variant-specific Wald estimates were calculated (β coefficients for variant-outcome [breast cancer risk] associations divided by β coefficients for variant-trait [risk for each of the other 16 cancers] associations) and combined as an estimate for the effect of the trait (each cancer PRS) on breast cancer risk using the IVW meta-analysis. Odds ratios (ORs) of breast cancer risk were estimated per unit of increase in PRS for each cancer. p for heterogeneity was estimated between ER-positive and ER-negative and across the five intrinsic subtypes using Cochran's Q statistic.

To replicate findings obtained from the BCAC, we analyzed individual-level data obtained from European ancestry women included in the UK Biobank cohort. Since some of the risk variants were not available in the UK Biobank, variants in high LD (n = 3; $r^2 \geq 0.85$) with these previously reported variants were selected as substitutes for the study. After imputation, no participant had a missing value for the 497 risk variants associated with the 16 types of cancer selected for this study (Table S3). We constructed a PRS for each of the 16 types of cancer using the same set of risk variants. Each

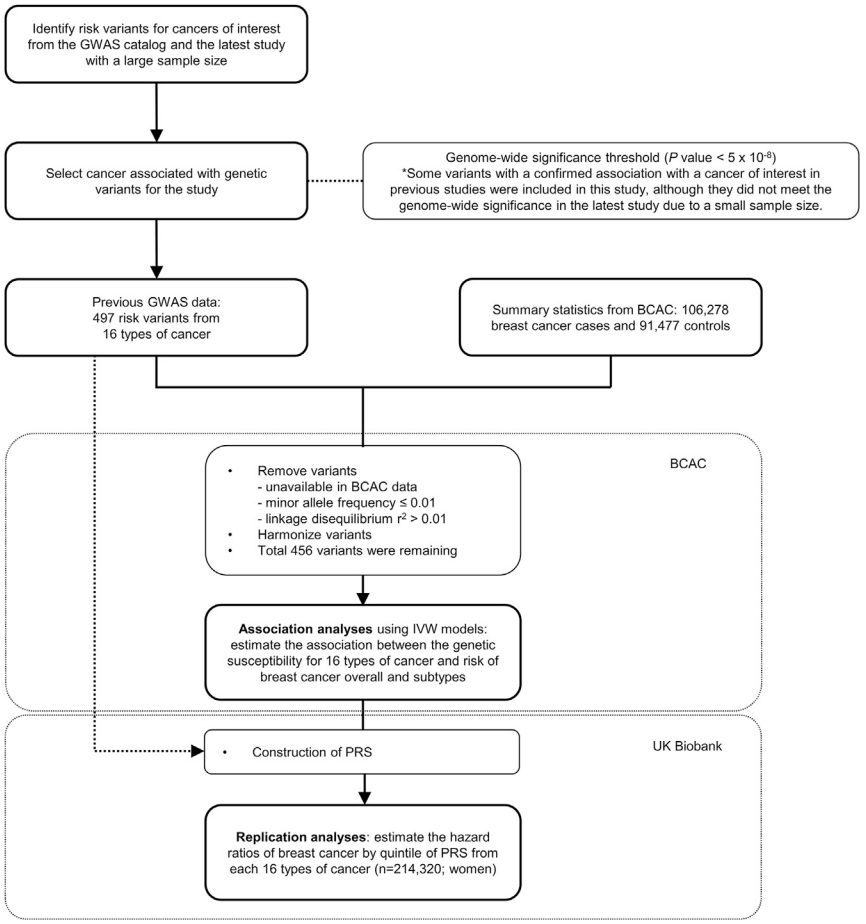

Figure 1. Flow chart for study design
BCAC, Breast Cancer Association Consortium; GWAS, genome-wide association study; IVW, inverse-variance weighted; PRS, polygenic risk score.

cancer-specific PRS was built using risk variants identified in previous GWAS for that cancer (Table S5). We calculated the PRS by summing the product of the weight (regression coefficient derived from previous GWAS) and the number of risk alleles (0, 1, and 2) for each risk variant across all GWAS-identified risk variants for that cancer. Details on the derivation of the genetic risk score have been published recently.[17,25,26] Hazard ratios (HRs) and 95% confidence intervals (CIs) associated with each PRS were estimated by Cox proportional hazard models using age as the underlying timescale left-truncated at the age of baseline interview and adjusted for age at enrollment, genotype array type (UKBL or UKBB), the 10 PCs for ancestry, and stratified by birth cohorts. The assumptions of proportionality were examined using Schoenfeld residuals.

Among 456 variants included in the IVW analysis of BCAC data, seven were associated with risk of more than one cancer (Table S4). Therefore, 449 variants remained for the analysis to evaluate the associations of breast cancer risk with each of these unique variants included in the 16 cancer-specific PRSs, with an adjustment of previously reported risk variants for breast cancer located within 1 Mb. We used genome-wide complex trait analysis (GCTA) software (option, COJO) to perform the conditional analyses.[14] GWAS summary statistics from BCAC were used in the analyses,[17,19] and individuals of European ancestry in the 1000 Genomes Project were used as the LD reference panel.

Statistical inferences were based on two-sided tests at a significance level of 0.05 unless otherwise specified using SAS software, v.9.4 (SAS Institute, Cary, NC) and R v.3.6.0 software.

Results

Associations of 16 cancer-specific PRSs with breast cancer risk: Results based on the BCAC dataset

We found a 5%–8% significantly elevated risk of overall breast cancer associated with the PRS for glioma or cancer of the corpus uteri, stomach, or colorectum based on the BCAC dataset. The positive associations with PRS of corpus uteri (OR = 1.09; 95% CI, 1.03–1.15; p = 0.002) or stomach (OR = 1.07; 95% CI, 1.03–1.12; p = 0.001) were statistically significant only for ER-positive cancer. On the other hand, the positive associations with the PRSs of colorectal and lung cancers were statistically significant only for ER-negative breast cancer risk (colorectum: OR = 1.08; 95% CI, 1.01–1.15; p = 0.033; lung: OR = 1.14; 95% CI, 1.00–1.30; p = 0.048) (Figure 2). Heterogeneity tests, however, were not statistically significant by ER status for any of the associations mentioned above.

Analyses by intrinsic subtype revealed associations for specific breast cancer subtypes, including a positive association of colorectal, corpus uteri, and stomach cancer PRSs with luminal A breast cancer risk (colorectum: OR = 1.06; 95% CI, 1.01–1.11; p = 0.018; corpus uteri: OR = 1.10; 95% CI, 1.03–1.17; p = 0.003; stomach: OR = 1.10; 95% CI, 1.05–1.15; p < 0.0001). We also found other subtype-specific associations, which included luminal B/HER2-negative breast cancer with the PRS for melanoma (OR = 1.11; 95% CI, 1.04–1.19; p = 0.004); HER2-enriched breast cancer with colorectal cancer PRS (OR = 1.13; 95% CI, 1.03–1.25; p = 0.013) and kidney cancer PRS (OR = 1.11; 95% CI, 1.01–1.23; p = 0.033); and triple-negative breast cancer with ovarian cancer PRS (OR = 1.25; 95% CI, 1.01–1.55; p = 0.040) (Figure 3; Table S7). The stomach cancer PRS was primarily associated with luminal A breast cancer (p for heterogeneity = 0.01), while melanoma PRS was limited to luminal B/HER2-negative breast cancer (p for heterogeneity = 0.05) (Figure 3).

Replication analyses using UK Biobank data

We used the data from the UK Biobank to replicate our findings (Figure 4). We compared the overall breast cancer risk among individuals with the top 10% PRS versus the

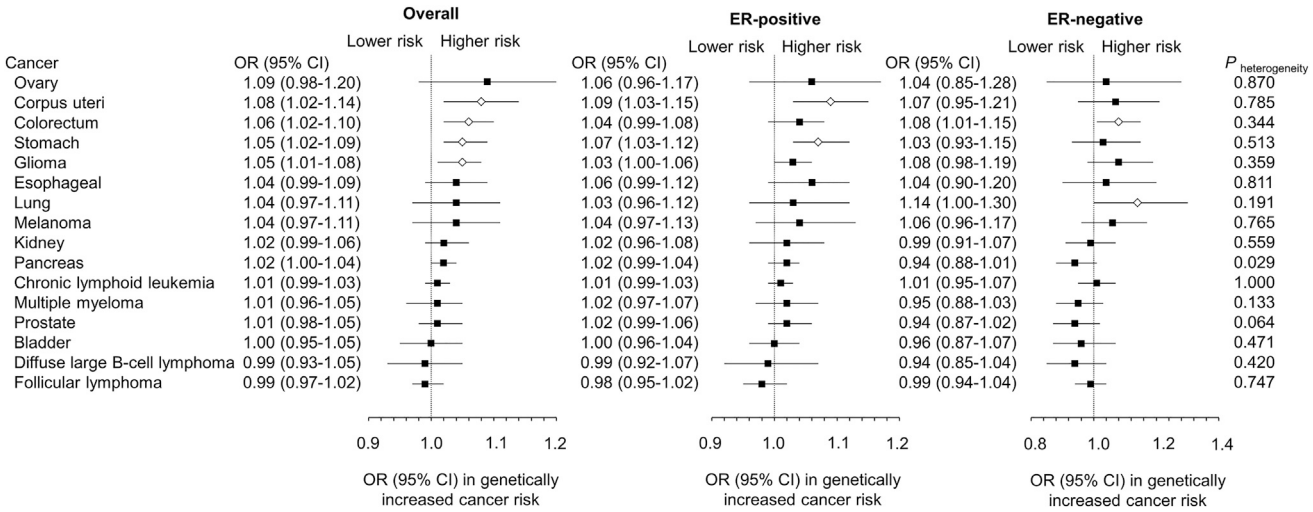

Figure 2. Association between the polygenic risk scores for 16 types of cancer and risk of breast cancer overall, ER-positive, and ER-negative

White rhombus indicates $p < 0.05$. Odds ratios of breast cancer risk were estimated per unit of increase in PRS for each cancer. p for heterogeneity was tested between ER-positive and ER-negative. CI, confidence interval; ER, estrogen receptor; OR, odds ratio.

bottom 10% PRS using the Cox proportional hazard regression models. The positive associations of overall breast cancer risk with PRS for cancers of the stomach, lung, colorectum, ovary, and melanoma were replicated using UK Biobank data, with an OR ranging from 1.14 to 1.19. The HR for overall breast cancer per standard deviation of PRS is also shown in Table S8.

Associations of individual variants with breast cancer risk

We further examined the association of individual variants included in the 16 cancer PRSs with overall breast cancer and its subtypes using BCAC data (Figure S1). Of the 449 unique risk variants, 45 variants were associated with overall breast cancer at $p < 1.11 \times 10^{-4}$ (0.05/449 variants, the Bonferroni corrected significance threshold) with a positive association direction found for 35 variants and negative association direction found for 10 variants (Figure S1A). Statistically significant associations were found for 26 variants with ER-positive cancer and 5 variants with ER-negative cancer (Figures S1B and S1C). In addition, we also found statistically significant associations for 28 variants with luminal A cancer, 3 variants with luminal B/HER2-negative cancer, 4 variants with luminal B cancer, 1 variant with HER2-enriched cancer, and 14 variants with triple-negative cancer (Figure S1D).

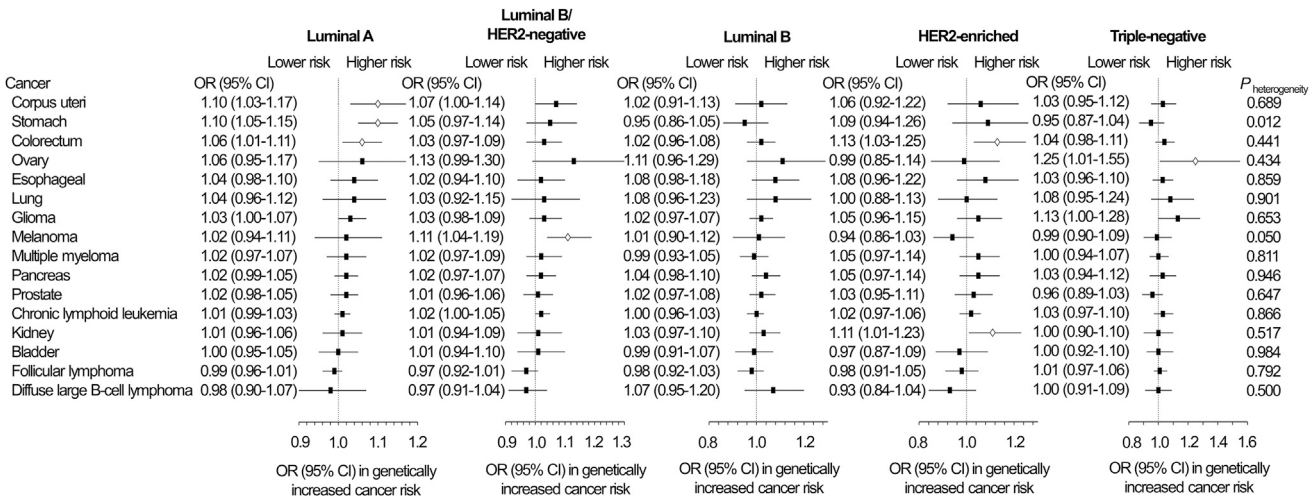

Figure 3. Association between the polygenic risk scores for 16 types of cancer and risk of intrinsic breast cancer subtypes

White rhombus indicates $p < 0.05$. Odds ratios of breast cancer risk were estimated per unit of increase in PRS for each cancer. p for heterogeneity was tested across five intrinsic breast cancer subtypes. CI, confidence interval; HER2, human epidermal growth factor receptor 2; OR, odds ratio.

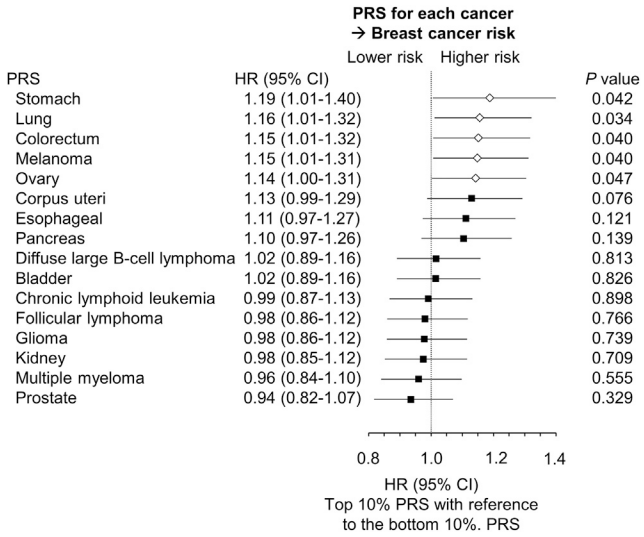

Figure 4. Hazard ratios of breast cancer associated with the top 10% versus bottom 10% of PRS from each of the 16 types of cancer, UK Biobank

White rhombus indicates $p < 0.05$. Hazard ratios were estimated using Cox regression and adjusted for age, genotyping array, and top 10 PCs for ancestry, and stratified by birth cohort. We performed all analyses only in women. CI, confidence interval; HR, hazards ratio; PRS, polygenic risk scores.

Conditional analyses adjusting for previously identified breast cancer risk variants within 1 Mb revealed seven variants not yet been reported in association with breast cancer risk previously (Table 1).

Discussion

This is the first large study to evaluate PRSs of major cancers in association with breast cancer risk. Our findings provided evidence that the PRS for glioma and cancers of the corpus uteri, stomach, and colorectum are associated with overall breast cancer risk. Of note, some PRS associations differed by subtypes of breast cancer, supporting the notion that there are some differences in genetic susceptibility across breast cancer subtypes.

Our findings are supported, in part, by previous studies regarding cancer genetic pleiotropy. A previous study of six solid cancers found moderate genetic correlations of breast cancer with both lung and colorectal cancers.[29] Furthermore, a recent study by the same research group with an increased sample size demonstrated that ovarian, colorectal, and lung cancers shared genetic susceptibility with breast cancer.[13] That study also observed a significantly higher genetic correlation of lung cancer with ER-negative than ER-positive breast cancer, which is consistent with the findings of our study. Another study found that breast cancer had a positive genetic correlation with bladder and esophageal/stomach cancers.[30] However, in a previous study that evaluated 13 different cancers, no significant genetic correlations between breast cancer and other cancers were observed, although a marginally significant genetic correlation was found for four cancer pairs.[12] Recently, Graff et al. evaluated potential pleiotropic effects of PRSs for 16 cancers and found a significant association between melanoma PRS and breast cancer risk (OR = 1.04; $p = 6.33 \times 10^{-7}$).[31] This finding is consistent with our result for a significant association of melanoma PRS with luminal B/HER2-negative breast cancer. Instead of estimating genetic correlation of breast cancer with other cancers, we used cancer-specific PRSs derived using risk variants identified from GWAS to quantify the risk of breast cancer in association with genetic susceptibility to other cancers, which provides additional insights into the genetics and etiology of breast and other cancers.

Table 1. Associations of breast cancer risk with seven genetic variants not previously reported in association with breast cancer risk

							Original GWAS (BCAC)		Conditional analysis		
SNP	Chr	Position	Nearest gene	Associated cancer	EA/RA	EAF	OR (95% CI)	p[a]	OR (95% CI)	p[a]	Associations
rs1321311	6	36622900	none	colorectum	A/C	0.240	1.03 (1.02–1.05)	1.20×10^{-5}	1.03 (1.02–1.04)	2.07×10^{-5}	overall
rs2811710	9	21991923	*CDKN2A*	multiple myeloma	C/T	0.641	0.95 (0.94–0.96)	3.69×10^{-15}	0.97 (0.96–0.98)	1.43×10^{-6}	overall
rs7931342	11	68994497	none	prostate	G/T	0.506	1.03 (1.01–1.04)	2.95×10^{-5}	1.03 (1.02–1.04)	2.00×10^{-6}	overall
rs11214775	11	113807181	*HTR3B*	prostate	G/A	0.712	1.03 (1.02–1.05)	1.04×10^{-6}	1.03 (1.02–1.05)	9.48×10^{-7}	overall
rs4924487	15	40922915	*KNL1*	prostate	C/G	0.841	1.04 (1.02–1.06)	1.60×10^{-6}	1.04 (1.02–1.06)	1.60×10^{-6}	overall
rs111572611	15	66755923	*MAP2K1*	prostate	T/C	0.257	1.03 (1.01–1.04)	3.81×10^{-5}	1.03 (1.02–1.04)	7.97×10^{-6}	overall
rs17601876	15	51553909	*CYP19A1*, *MIR4713HG*	corpus uteri	G/A	0.480	1.03 (1.02–1.05)	5.47×10^{-5}	1.03 (1.02–1.05)	5.47×10^{-5}	luminal A

BCAC, Breast Cancer Association Consortium; Chr, chromosome; CI, confidence interval; EA/RA, effective allele/reference allele; EAF, effective allele frequency; OR, odds ratio.

[a]Significance threshold was set at $p < 1.1 \times 10^{-4}$ with the adjustment of 449 variants evaluated in this analysis.

A unique strength of our study is the ability to evaluate genetic associations according to breast cancer subtypes. Cumulative evidence supports the notion that there are some differences in the etiology across breast cancer by subtypes. For example, ER-negative cancer shows a weaker association with reproductive risk factors than ER-positive cancer.[32,33] In our study, we found that colorectal and lung cancer PRSs showed significant associations with ER-negative but not ER-positive breast cancer, although heterogeneity test was not statistically significant. Furthermore, we observed a significant association between ovarian cancer PRS and triple-negative breast cancer. It is well established that *BRCA1* (MIM: 113705) pathogenetic mutation carriers have a high risk of ovarian cancer and are more likely to develop triple-negative or basal-like breast cancer than other types.[34–36] In addition, other germline pathogenic mutations were also detected in triple-negative breast cancer patients, with the majority observed in genes involved in homologous recombination, including *PALB2* (MIM: 610355), *BARD1* (MIM: 601593), *RAD51C* (MIM: 602774), and *RAD51D* (MIM: 602954).[37–40] The deleterious germline mutations of these genes are also associated with ovarian cancer risk.[41,42] Our study expanded the knowledge regarding shared rare pathogenetic germline mutations between triple-negative breast cancer and ovarian cancer and suggests that these two cancers may also share certain common genetic risk variants.

Rather than a genome-wide search, we conducted a focused study evaluating 449 unique variants included in the 16 cancer-specific PRSs. With a reduced number of comparisons, we identified seven variants associated with breast cancer risk that had not been reported from previous studies. There is some biological evidence to support the association of breast cancer risk with these newly identified genetic variants. For example, several studies suggested that cyclin-dependent kinase inhibitor 2A (*CDKN2A* gene, encoding for tumor suppressor proteins, MIM: 600160), well known as a susceptibility gene for melanoma and pancreatic cancer, may also be involved in breast tumorigenesis.[43–45] Activation of mitogen-activated protein kinases (*MAPK*) in breast cancer leads to increased proliferation, invasion, and metastasis of breast cancer.[46–48] The *CYP19A1* (MIM: 107910) gene encodes the enzyme responsible for biosynthesis of estrogen, a sex hormone that plays a central role in the etiology of breast cancer.[49–52] In our study, one variant (rs17601876 nearest to *CYP19A1* gene) showed a significant association with luminal A breast cancer. These findings need to be replicated in future studies with a larger sample size.

Our study systematically collected the most recent GWAS-identified risk variants and used the data from the BCAC, a consortium using a case-control study design, and the UK Biobank, with a cohort study design. Both studies have a very large sample size, and the findings from these two datasets supported each other. We used IVW meta-analyses, a well-accepted two-sample analytic approach in Mendelian randomization analyses using summary-level data. Our results support the hypothesis that overall genetic susceptibility to certain cancers may be causally related to breast cancer risk. We replicated our results from the BCAC using individual-level data from the UK Biobank. However, as the number of breast cancer cases in the UK Biobank is relatively small, we could not evaluate the associations by breast cancer subtypes.

In conclusion, our study provides strong evidence that there is shared genetic susceptibility between breast cancer and several other cancers. The shared genetic susceptibility may differ by breast cancer subtype. Results from our study help uncover the genetic basis for breast and other cancers and identify individuals at high risk for multiple cancers.

Data and code availability

The summary statistics data from the BCAC used in this study are available in its study website. The individual data from the UK Biobank used in this project, along with code books and codes, can be obtained directly from the UK Biobank by submitting a data request proposal.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2021.100077.

Acknowledgments

This work was supported in part by the Anne Potter Wilson chair endowment at Vanderbilt University as well as NIH grants R01 CA202981 and R01 CA235553. This research has been conducted using the UK Biobank Resource under application no. 40685 and BCAC data. We acknowledge all the families and clinicians who contributed to the participating studies.

Declaration of interests

The authors declare no competing interests.

Received: August 16, 2021
Accepted: December 6, 2021

Web resources

Breast Cancer Association Consortium, http://bcac.ccge.medschl.cam.ac.uk/
UK Biobank, https://www.ukbiobank.ac.uk/
GWAS catalog, https://www.ebi.ac.uk/gwas/
R v.3.6.0 software, https://www.r-project.org/
Online Mendelian Inheritance in Man, http://www.omim.org.

References

1. Cheng, I., Kocarnik, J.M., Dumitrescu, L., Lindor, N.M., Chang-Claude, J., Avery, C.L., Caberto, C.P., Love, S.-A., Slattery, M.L., Chan, A.T., et al. (2014). Pleiotropic effects of

genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia. Gut *63*, 800–807.
2. Crowther-Swanepoel, D., Broderick, P., Di Bernardo, M.C., Dobbins, S.E., Torres, M., Mansouri, M., Ruiz-Ponte, C., Enjuanes, A., Rosenquist, R., Carracedo, A., et al. (2010). Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet *42*, 132–136.
3. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature *447*, 1087–1093.
4. Rothman, N., Garcia-Closas, M., Chatterjee, N., Malats, N., Wu, X., Figueroa, J.D., Real, F.X., Van Den Berg, D., Matullo, G., Baris, D., et al. (2010). A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat Genet *42*, 978–984.
5. Shete, S., Hosking, F.J., Robertson, L.B., Dobbins, S.E., Sanson, M., Malmer, B., Simon, M., Marie, Y., Boisselier, B., Delattre, J.-Y., et al. (2009). Genome-wide association study identifies five susceptibility loci for glioma. Nat Genet *41*, 899–904.
6. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet *40*, 310–315.
7. Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet *39*, 645–649.
8. Hu, Z., Wu, C., Shi, Y., Guo, H., Zhao, X., Yin, Z., Yang, L., Dai, J., Hu, L., Tan, W., et al. (2011). A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat Genet *43*, 792–796.
9. Haiman, C.A., Chen, G.K., Vachon, C.M., Canzian, F., Dunning, A., Millikan, R.C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C.B., et al. (2011). A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet *43*, 1210–1214.
10. Peters, U., Hutter, C.M., Hsu, L., Schumacher, F.R., Conti, D.V., Carlson, C.S., Edlund, C.K., Haile, R.W., Gallinger, S., Zanke, B.W., et al. (2012). Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet *131*, 217–234.
11. Turnbull, C., Rapley, E.A., Seal, S., Pernet, D., Renwick, A., Hughes, D., Ricketts, M., Linger, R., Nsengimana, J., Deloukas, P., et al. (2010). Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet *42*, 604–607.
12. Sampson, J.N., Wheeler, W.A., Yeager, M., Panagiotou, O., Wang, Z., Berndt, S.I., Lan, Q., Abnet, C.C., Amundadottir, L.T., Figueroa, J.D., et al. (2015). Analysis of heritability and shared heritability based on genome-wide association studies for thirteen cancer types. J Natl Cancer Inst *107*, djv279.
13. Jiang, X., Finucane, H.K., Schumacher, F.R., Schmit, S.L., Tyrer, J.P., Han, Y., Michailidou, K., Lesseur, C., Kuchenbaecker, K.B., Dennis, J., et al. (2019). Shared heritability and functional enrichment across six solid cancers. Nat Commun *10*, 1–23.
14. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet *88*, 76–82.
15. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen, C., Duncan, L., Perry, J.R., Patterson, N., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet *47*, 1236–1241.
16. Holm, J., Eriksson, L., Ploner, A., Eriksson, M., Rantalainen, M., Li, J., Hall, P., and Czene, K. (2017). Assessment of breast cancer risk factors reveals subtype heterogeneity. Cancer Res *77*, 3708–3717.
17. Zhang, H., Ahearn, T.U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., Jiang, X., O'Mara, T.A., Zhao, N., Bolla, M.K., et al. (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet *52*, 572–581.
18. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med *12*, e1001779.
19. Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. Nature *551*, 92–94.
20. Milne, R.L., Kuchenbaecker, K.B., Michailidou, K., Beesley, J., Kar, S., Lindström, S., Hui, S., Lemaçon, A., Soucy, P., Dennis, J., et al. (2017). Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet *49*, 1767–1778.
21. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K., et al. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet *45*, 353–361.
22. Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M.J., Maranian, M.J., Bolla, M.K., Wang, Q., Shah, M., et al. (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet *47*, 373–380.
23. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.
24. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.
25. Jia, G., Lu, Y., Wen, W., Long, J., Liu, Y., Tao, R., Li, B., Denny, J.C., Shu, X.O., and Zheng, W. (2020). Evaluating the utility of polygenic risk scores in identifying high-risk individuals for eight common cancers. JNCI Cancer Spectr *4*, pkaa021.
26. Choi, J., Jia, G., Wen, W., Long, J., and Zheng, W. (2020). Evaluating polygenic risk scores in assessing risk of nine solid and hematologic cancers in European descendants. Int J Cancer *147*, 3416–3423.
27. Borenstein, M., Hedges, L.V., Higgins, J.P., and Rothstein, H.R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods *1*, 97–111.
28. Burgess, S., Davey Smith, G., Davies, N.M., Dudbridge, F., Gill, D., Glymour, M.M., Hartwig, F.P., Holmes, M.V., Minelli, C., Relton, C.L., and Theodoratou, E. (2020). Guidelines for

performing Mendelian randomization investigations. Wellcome Open Res *4*, 186.
29. Lindström, S., Finucane, H., Bulik-Sullivan, B., Schumacher, F.R., Amos, C.I., Hung, R.J., Rand, K., Gruber, S.B., Conti, D., Permuth, J.B., et al. (2017). Quantifying the genetic correlation between multiple cancer types. Cancer Epidemiol Biomarkers Prev *26*, 1427–1435.
30. Rashkin, S.R., Graff, R.E., Kachuri, L., Thai, K.K., Alexeeff, S.E., Blatchins, M.A., Cavazos, T.B., Corley, D.A., Emami, N.C., Hoffman, J.D., et al. (2020). Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. Nat Commun *11*, 1–14.
31. Graff, R.E., Cavazos, T.B., Thai, K.K., Kachuri, L., Rashkin, S.R., Hoffman, J.D., et al. (2021). Cross-cancer evaluation of polygenic risk scores for 16 cancer types in two large cohorts. Nat Commun *12*, 970.
32. Blows, F.M., Driver, K.E., Schmidt, M.K., Broeks, A., van Leeuwen, F.E., Wesseling, J., Cheang, M.C., Gelmon, K., Nielsen, T.O., Blomqvist, C., et al. (2010). Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med *7*, e1000279.
33. Yang, X.R., Chang-Claude, J., Goode, E.L., Couch, F.J., Nevanlinna, H., Milne, R.L., Gaudet, M., Schmidt, M.K., Broeks, A., Cox, A., et al. (2011). Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst *103*, 250–263.
34. Schorge, J.O., Modesitt, S.C., Coleman, R.L., Cohn, D.E., Kauff, N.D., Duska, L.R., and Herzog, T.J. (2010). SGO White Paper on ovarian cancer: etiology, screening and surveillance. Gynecol Oncol *119*, 7–17.
35. Tschernichovsky, R., and Goodman, A. (2017). Risk-reducing strategies for ovarian cancer in BRCA mutation carriers: a balancing act. Oncologist *22*, 450–459.
36. Atchley, D.P., Albarracin, C.T., Lopez, A., Valero, V., Amos, C.I., Gonzalez-Angulo, A.M., Hortobagyi, G.N., and Arun, B.K. (2008). Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. J Clin Oncol *26*, 4282–4288.
37. Hu, C., Hart, S.N., Gnanaolivu, R., Huang, H., Lee, K.Y., Na, J., et al. (2021). A population-based study of genes previously implicated in breast cancer. N Engl J Med *384*, 440–451.
38. Breast Cancer Association, C., Dorling, L., Carvalho, S., Allen, J., Gonzalez-Neira, A., Luccarini, C., Wahlstrom, C., Pooley, K.A., Parsons, M.T., Fortuno, C., et al. (2021). Breast cancer risk genes—association analysis in more than 113,000 women. N Engl J Med *384*, 428–439.
39. Shimelis, H., LaDuca, H., Hu, C., Hart, S.N., Na, J., Thomas, A., Akinhanmi, M., Moore, R.M., Brauch, H., Cox, A., et al. (2018). Triple-negative breast cancer risk genes identified by multigene hereditary cancer panel testing. J Natl Cancer Inst *110*, 855–862.
40. Couch, F.J., Hart, S.N., Sharma, P., Toland, A.E., Wang, X., Miron, P., Olson, J.E., Godwin, A.K., Pankratz, V.S., Olswold, C., et al. (2015). Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. J Clin Oncol *33*, 304–311.
41. Ramus, S.J., Song, H., Dicks, E., Tyrer, J.P., Rosenthal, A.N., Intermaggio, M.P., et al. (2015). Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. J Natl Cancer Inst *107* (11), djv214.
42. Norquist, B.M., Harrell, M.I., Brady, M.F., Walsh, T., Lee, M.K., Gulsuner, S., Bernards, S.S., Casadei, S., Yi, Q., Burger, R.A., et al. (2016). Inherited mutations in women with ovarian carcinoma. JAMA Oncol *2*, 482–490.
43. Aftab, A., Shahzad, S., Hussain, H.M.J., Khan, R., Irum, S., and Tabassum, S. (2019). CDKN2A/P16INK4A variants association with breast cancer and their in-silico analysis. Breast Cancer *26*, 11–28.
44. Chan, S.H., Chiang, J., and Ngeow, J. (2021). CDKN2A germline alterations and the relevance of genotype-phenotype associations in cancer predisposition. Hered Cancer Clin Pract *19*, 21.
45. Potrony, M., Puig-Butille, J.A., Aguilera, P., Badenas, C., Carrera, C., Malvehy, J., and Puig, S. (2014). Increased prevalence of lung, breast, and pancreatic cancers in addition to melanoma risk in families bearing the cyclin-dependent kinase inhibitor 2A mutation: implications for genetic counseling. J Am Acad Dermatol *71*, 888–895.
46. Seger, R., and Krebs, E.G. (1995). The MAPK signaling cascade. FASEB J *9*, 726–735.
47. Blenis, J. (1993). Signal transduction via the MAP kinases: proceed at your own RSK. Proc Natl Acad Sci U S A *90*, 5889–5892.
48. Esteva, F.J., Sahin, A.A., Smith, T.L., Yang, Y., Pusztai, L., Nahta, R., Buchholz, T.A., Buzdar, A.U., Hortobagyi, G.N., and Bacus, S.S. (2004). Prognostic significance of phosphorylated P38 mitogen-activated protein kinase and HER-2 expression in lymph node-positive breast carcinoma. Cancer *100*, 499–506.
49. Dunning, A.M., Dowsett, M., Healey, C.S., Tee, L., Luben, R.N., Folkerd, E., Novik, K.L., Kelemen, L., Ogata, S., Pharoah, P.D., et al. (2004). Polymorphisms associated with circulating sex hormone levels in postmenopausal women. J Natl Cancer Inst *96*, 936–945.
50. Flote, V.G., Furberg, A.S., McTiernan, A., Frydenberg, H., Ursin, G., Iversen, A., Lofteroed, T., Ellison, P.T., Wist, E.A., Egeland, T., et al. (2014). Gene variations in oestrogen pathways, CYP19A1, daily 17beta-estradiol and mammographic density phenotypes in premenopausal women. Breast Cancer Res *16*, 499.
51. Simpson, E., and Santen, R.J. (2015). Celebrating 75 years of oestradiol. J Mol Endocrinol *55*, T1–T20.
52. Zhao, H., Zhou, L., Shangguan, A.J., and Bulun, S.E. (2016). Aromatase expression and regulation in breast and endometrial cancer. J Mol Endocrinol *57*, R19–R33.