CANCER RESEARCH COMMUNICATIONS

AACR American Association for Cancer Research

RESEARCH ARTICLE https://doi.org/10.1158/2767-9764.CRC-21-0119 OPEN ACCESS

A Genome-Wide Gene-Based Gene–Environment Interaction Study of Breast Cancer in More than 90,000 Women

Xiaoliang Wang[1,2], Hongjie Chen[1], Pooja Middha Kapoor[3,4], Yu-Ru Su[2], Manjeet K. Bolla[5], Joe Dennis[5], Alison M. Dunning[6], Michael Lush[5], Qin Wang[5], Kyriaki Michailidou[7,8], Paul D.P. Pharoah[5,6], John L. Hopper[9], Melissa C. Southey[10,11,12], Stella Koutros[13], Laura E. Beane Freeman[13], Jennifer Stone[14], Gad Rennert[15], Rana Shibli[15], Rachel A. Murphy[16], Kristan Aronson[17], Pascal Guénel[18], Thérèse Truong[18], Lauren R. Teras[19], James M. Hodge[19], Federico Canzian[20], Rudolf Kaaks[20], Hermann Brenner[21], Volker Arndt[21], Reiner Hoppe[22,23], Wing-Yee Lo[22,23], Sabine Behrens[3], Arto Mannermaa[24,25,26], Veli-Matti Kosma[24,25,26], Audrey Jung[3], Heiko Becher[27], Graham G. Giles[9,10,11], Christopher A. Haiman[28], Gertraud Maskarinec[29], Christopher Scott[30], Stacey Winham[30], Jacques Simard[31], Mark S. Goldberg[32], Wei Zheng[33], Jirong Long[33], Melissa A. Troester[34], Michael I. Love[34], Cheng Peng[35], Rulla Tamimi[36], Heather Eliassen[37], Montserrat García-Closas[13], Jonine Figueroa[38], Thomas Ahearn[13], Rose Yang[13], D. Gareth Evans[39,40,41], Anthony Howell[42], Per Hall[43], Kamila Czene[43], Alicja Wolk[44], Dale P. Sandler[45], Jack A. Taylor[45], Anthony J. Swerdlow[46,47], Nick Orr[48], James V. Lacey[49], Sophia Wang[49], Håkan Olsson[50,†], Douglas F. Easton[5,6], Roger L. Milne[9,10,11], Li Hsu[2,51], Peter Kraft[37,52], Jenny Chang-Claude[3,53], and Sara Lindström[1,2]

ABSTRACT

Genome-wide association studies (GWAS) have identified more than 200 susceptibility loci for breast cancer, but these variants explain less than a fifth of the disease risk. Although gene–environment interactions have been proposed to account for some of the remaining heritability, few studies have empirically assessed this.

We obtained genotype and risk factor data from 46,060 cases and 47,929 controls of European ancestry from population-based studies within the Breast Cancer Association Consortium (BCAC). We built gene expression prediction models for 4,864 genes with a significant ($P < 0.01$) heritable component using the transcriptome and genotype data from the Genotype-Tissue Expression (GTEx) project. We leveraged predicted gene expression information to investigate the interactions between gene-centric genetic variation and 14 established risk factors in association with breast cancer risk, using a mixed-effects score test.

After adjusting for number of tests using Bonferroni correction, no interaction remained statistically significant. The strongest interaction observed was between the predicted expression of the *C13orf45* gene and age at first full-term pregnancy ($P_{GXE} = 4.44 \times 10^{-6}$).

In this transcriptome-informed genome-wide gene–environment interaction study of breast cancer, we found no strong support for the role of gene expression in modifying the associations between established risk factors and breast cancer risk.

Our study suggests a limited role of gene–environment interactions in breast cancer risk.

[1]Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington. [2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington. [3]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [4]Faculty of Medicine, University of Heidelberg, Heidelberg, Germany. [5]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom. [6]Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom. [7]Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. [8]Cyprus School of Molecular Medicine, Nicosia, Cyprus. [9]Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia. [10]Precision Medicine, School of Clinical Sciences at

Introduction

Breast cancer is the most commonly diagnosed malignancy in women. In 2020, breast cancer was estimated to be newly diagnosed in 2.3 million women, and meanwhile caused more than 680,000 deaths worldwide (1). Both genetic and environmental factors have been found to contribute to the etiology of breast cancer. Twin studies have estimated that approximately 30% of variance in breast cancer incidence can be explained by genetic variation (2, 3). Genome-wide association studies (GWAS) have identified more than 200 independent loci that are associated with breast cancer risk (4). However, these single-nucleotide polymorphisms (SNPs) only explain approximately 19% of the familial relative risk. Meanwhile, observational studies have demonstrated that several environmental and lifestyle risk factors, including age at menarche, body mass index (BMI), alcohol consumption, parity, and use of menopausal hormone therapy (MHT), also affect the risk of breast cancer (5–11). Exploring the interplay of genetic and environmental risk factors (GxE interactions) is thus crucial in understanding the development of breast cancer.

The Breast Cancer Association Consortium (BCAC) has published multiple studies which reported various interactions between individual SNPs and established risk factors. Nickels and colleagues reported potential interactions between genetic variants and several environmental and lifestyle factors, including number of full-term pregnancies, alcohol consumption, and ever being parous (12). Schoeps and colleagues reported that two SNPs on locus *21q22.12* may interact with postmenopausal BMI to significantly affect the risk of breast cancer (13). However, other previous genome-wide gene–environmental interaction studies (GWEIS) reported no statistically significant interactions between SNPs and established breast cancer risk factors (4, 14–20). Statistical power remains one of the primary issues in GWEIS, as they require much larger sample sizes for detecting interactions as compared with marginal associations of similar magnitude (4, 21).

Novel statistical methods, such as gene-based testing that incorporates functional information, can substantially reduce the burden of multiple comparisons. As most GWAS hits fall outside of the coding region of genes and are enriched in regulatory elements, it has been hypothesized that many GWAS-identified genotype–phenotype associations are driven by the regulatory function on the expression of nearby genes (22–24). Wu and colleagues conducted a transcriptome-wide association study (TWAS) of breast cancer that systematically investigated the association between predicted gene expression and disease risk, and reported 48 statistically significant genes associations (25). These results suggest that incorporating SNP-specific regulatory information on gene expression could help discovering meaningful GxE interactions.

In this study, we utilized the genotype and environmental risk factor data collected by the Breast Cancer Association Consortium (BCAC). Using breast tissue–specific transcriptome and genotype data from the Genotype-Tissue Expression (GTEx) project, we built gene expression prediction models for 4,864 genes with a significant heritable component. We then systematically assessed the interactions between these genes and 14 established risk factors in relation to the risk of breast cancer, using a mixed-effects score test called MiSTi (mixed-effects score test for interactions; ref. 26). Our study is the first to incorporate genetically determined gene expression data in the investigation of GxE interactions in breast cancer.

Monash Health, Monash University, Clayton, Victoria, Australia. [11]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia. [12]Department of Clinical Pathology, The University of Melbourne, Victoria, Australia. [13]Division of Cancer Epidemiology and Genetic, NCI, NIH, Bethesda, Maryland. [14]Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Crawley, Australia. [15]Department of Community Medicine and Epidemiology, Carmel Medical Center, Haifa, Israel. [16]Cancer Control Research, BC Cancer and School of Population and Public Health, University of British Columbia, Vancouver, Canada. [17]Public Health Sciences, Queen's University, Kingston, Canada. [18]Université Paris-Saclay, Inserm, CESP, Team Exposome and Heredity, Villejuif, France. [19]Department of Population Science, American Cancer Society, Atlanta, Georgia. [20]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. [21]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [22]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany. [23]University of Tübingen, Tübingen, German. [24]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland. [25]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland. [26]Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland. [27]Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [28]Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California. [29]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii. [30]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota. [31]Genomics Center, Centre Hospitalier Universitaire de Québec-Université Laval Research Center, Québec City, Quebec, Canada. [32]Department of Medicine, McGill University, Montréal, Quebec, Canada; Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, Quebec, Canada. [33]Division of Epidemiology, Vanderbilt University Medical Center, Nashville, Tennessee. [34]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina. [35]Channing Division of Network Medicine, Department of Medicine, Brigham & Women's Hospital, Boston, Massachusetts. [36]Department of Population Health Sciences, Weill Cornell Medicine, New York, New York. [37]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [38]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh Medical School, Edinburgh, United Kingdom. [39]Division of Evolution and Genomic Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom. [40]Genomic Medicine, St Mary's Hospital, Manchester Centre for Genomic Medicine, Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom. [41]NIHR Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust, Manchester, United Kingdom. [42]Division of Cancer Sciences, University of Manchester, Manchester, United Kingdom. [43]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [44]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [45]Epidemiology Branch, Division of Intramural Research, National Institute of Environmental Health Sciences, National Institute of Health, Research Triangle Park, North Carolina. [46]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom. [47]Division of Breast Cancer Research, The Institute of Cancer Research, London, United K.ingdom [48]Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, United Kingdom. [49]Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, California. [50]Departments of Oncology and Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden. [51]Department of Biostatistics, School of Public Health, University of Washington, Seattle, Washington. [52]Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [53]Cancer Epidemiology Group, University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCCH), Hamburg, Germany.

X. Wang and H. Chen contributed equally to this article.

[†]Deceased.

Corresponding Author: Sara Lindström, Department of Epidemiology, School of Public Health, University of Washington, 3980 15th Avenue NE, Box 351619, Seattle, WA 98195. Phone: 206-221-3148; E-mail: saralind@uw.edu

doi: 10.1158/2767-9764.CRC-21-0119

Materials and Methods

Study Sample

For this study, we obtained breast cancer cases and controls from the cohort studies and population-based case–control studies participating in BCAC. BCAC is a well-established, international collaborative consortium of 84 epidemiologic and clinical breast cancer studies, which is integrated by investigators interested in the inherited risk of breast cancer (4). Genotype data were generated using either the iCOGs or OncoArray genotyping platforms. Both SNP arrays were customized and manufactured by Illumina, and consisted of 211,155 (iCOGs) and 533,000 (OncoArray) SNPs, respectively. In total, our study included 93,989 women (73,441 genotyped by OncoArray, 20,548 genotyped by iCOGS) from 31 studies, including ABCFS (27), AHS (28), BCEES (29), BCINIS (30, 31), CBCS (32–35), CECILE (36), CPSII (37), CTS (38), EPIC (39), ESTHER (40), GENICA (41, 42), GESBC (43), KARMA (44), KBCP (45, 46), MARIE (47), MCCS (48), MEC (49), MISS (50, 51), MMHS (52), MTLGEBCS (53), NBHS (54), NCBCS (55, 56), NHS (57), NHS2 (58), PBCS (59), PLCO (60), PROCAS (61), SASBAC (62), SISTER (63, 64), SMC (65), and UKBGS (ref. 66; Supplementary Table S1). In total, our study included 46,060 breast cancer cases (35,561 genotyped by OncoArray, 10,499 genotyped by iCOGs) and 47,929 controls (37,880 genotyped by OncoArray, 10,049 genotyped by iCOGs). All the women included were of European ancestry.

Details of the genotype calling, imputation, and quality control processes have been described elsewhere (67). Genotypes were imputed for all samples using the October 2014 (version 3) release of the 1000 Genomes Project dataset as the reference panel. The imputation was conducted using a two-stage approach, using SHAPEIT2 for phasing and IMPUTEv2 for imputation. Approximately 11.8 million SNPs with minor allele frequency (MAF) $>$ 0.5% and imputation quality score (INFO) $>$ 0.3 were included in our analysis.

Building the Prediction Model of Gene Expression

We used the RNA-sequencing and genotype data from 251 individuals published by the GTEx project version 7 to construct prediction models of gene expression in mammary tissue. Details of the GTEx project have been described elsewhere (68).

We built gene expression prediction models for each gene using the "FUSION" pipeline. Only the 1,217,312 SNPs included in the HapMap Phase 3 were included in building the prediction models. To estimate the genetically modulated expression of each gene, we included variants located within 500 kb on either side of the gene boundary. SNP-heritability of each gene was estimated using the REML algorithm implemented in the GCTA software (69). Gene expression models were constructed only if the SNP-heritability of gene expression was statistically significant at $P < 0.01$. Three prediction schemes, single best eQTL (Top1), LASSO regression, and elastic-net regression, were then utilized to build expression models for each heritable gene. The prediction accuracy of each derived model was then estimated using 5-fold cross-validation, and the best performing model was selected as the final model for each gene. We built gene expression prediction models for a total of 5,043 genes, of which we had breast cancer genotype data for 4,864 genes. The gene expression prediction models were then used as functional weights in the subsequent interaction analyses.

Collection of Breast Cancer Risk Factors

All demographic and breast cancer risk factor data were self-reported via interview or questionnaire prior to or shortly after breast cancer diagnosis (for cases) or the reference date (for controls, defined as the diagnosis date of matched breast cancer case). A total of 14 risk factors were included in the present analysis: age at first full-term pregnancy (per 5-year), average lifetime alcohol consumption (per 10 g/day), age at menarche (per 2-year), premenopausal BMI (per 5 kg/m^2), postmenopausal BMI (per 5 kg/m^2), breastfeeding history (yes/no), duration of breastfeeding (per 12-month), height (per 5 cm), history of oral contraceptive (OC) use (yes/no), parous (yes/no), number of full-term births (1/2/3/4+), current smoking status, current use of estrogen only (E-only) MHT, and current use of estrogen plus progestogen (E+P) MHT. BMI was analyzed separately for pre- and postmenopausal women, as the association between BMI and breast cancer risk varies across life stages (70). Analyses of reproductive factors were limited to parous women only and analyses of MHT use were limited to postmenopausal women.

Investigating Interactions between Predicted Gene Expression and Environmental Factors

We utilized a mixed-effects based analysis tool "MiSTi" (mixed-effects score test for interactions) to assess potential GxE interactions (26). MiSTi is a hierarchical model that assesses the joint interactions of a set of variants with environmental factors, by leveraging functional information across the variants. The GxE interaction is modeled by two components, one fixed and one random effects component. The fixed-effect component incorporates variant-specific functional information as weights to calculate the weighted burden of the variants, and then quantifies their interaction with the environmental factor. The random effects component involves any residual GxE interaction effect that cannot be addressed by the fixed effects. Here, the fixed effect component represents the interaction between predicted gene expression and the environmental factor, whereas the random effects component represents the residual interaction effects of any SNPs that were not accounted for in predicted gene expression. MiSTi includes a novel testing procedure, which derives two independent score statistics for the fixed effect and the random variance component separately and combines these two statistics through an adaptive weighted linear combination (aMiSTi) to assess the evidence of overall GxE interactions. The statistical power for GxE interaction analysis using MiSTi may be affected by multiple factors, including the LD structure of the gene, proportion of the variation in gene expression explained by the genetic regulatory variants, consistency of direction of effect between random and fixed effect, etc (71). Simulation analysis suggested that under type I error rate of 0.05, a sample size of 5,000 cases and 5,000 controls, for a gene harboring 100 genetic variants of which 27 were functional, MiTIi had an 81.3% of power to detect a significant GxE interaction using the aMiSTi approach when the fixed and random component had the same direction of interaction effect (26).

In each GxE interaction model, we adjusted for study, age (at diagnosis for cases; at reference date for controls), and first five principal components for population structure. For tests of current MHT use (E-only and E+P), we further adjusted for former use of the corresponding MHT (yes/no) in the model, to account for the association between former use of MHT (which attenuates with time since cessation) and breast cancer. To adjust for multiple comparisons, we considered any interactions with aMiSTi p-value $< 0.05/(4{,}864 \times 14) = 7.34 \times 10^{-7}$ as statistically significant. Because Bonferroni correction makes the strong assumption of independent tests and results in a stringent threshold for

TABLE 1 Distribution of environmental variables in the study population.

Continuous variables

	Cases		Controls	
Variable name	**Sample size**	**Mean (SD)**	**Sample size**	**Mean (SD)**
Age at menarche, y	43,138	12.91 (1.55)	45,513	12.99 (1.56)
Age at first full-term pregnancy[a], y	35,419	24.98 (4.67)	39,038	24.68 (4.55)
Duration of breastfeeding[a], mo	20,425	8.34 (10.96)	18,853	8,97 (11.35)
Adult BMI, Premenopausal[b], kg/m[b]	11,420	25.57 (5.47)	11,940	25.41 (5.17)
Adult BMI, Postmenopausal[c], kg/m[b]	31,036	26.78 (5.32)	33,213	26.39 (5.08)
Adult Height, cm	41,819	163.79 (6.45)	45,073	163.76 (6.45)
Lifetime alcohol consumption, g/day	22,653	6.53 (12.36)	21,337	5.72 (10.50)

Categorical variables

	Cases		Controls	
Variable name	**Sample size**	**%**	**Sample size**	**%**
Parity	43,465		45,771	
Parous	37,315	85.9	40,394	88.3
Nulliparous	6,150	14.1	5,377	11.7
Number of full-term births[a]	36,906		40,188	
1	6,714	18.2	6,147	15.3
2	15,578	42.2	16,966	42.2
3	8,910	24.1	10,061	25.0
4+	5,704	15.5	7,014	17.5
Ever breastfed[a]	25,135		23,561	
Yes	19,491	77.5	18,532	78.7
No	5,644	22.5	5,029	21.3
Ever use of OCs	41,359		43,269	
Yes	23,905	57.8	25,825	59.7
No	17,454	42.2	17,444	40.3
Smoking status	39,340		41,804	
Current	5,674	14.4	5,746	13.8
Former	12,136	30.9	12,845	30.7
Never	21,530	54.7	23,213	55.5
MHT use, Estrogen + Progestogen[c]	17,128		16,904	
Current	3,159	18.4	2,139	12.6
Former	1,557	9.1	1,554	9.2
Never	12,412	72.5	13,211	78.2
MHT Use, Estrogen[c]	17,163		16,911	
Current	2,685	15.6	2,855	16.9
Former	2,221	12.9	2,124	12.6
Never	12,257	71.5	11,932	70.5

[a]Among women with at least one full-term birth only.

[b]Among premenopausal women only.

[c]Among postmenopausal women only.

significance, we also report GxE interactions with a *P* value corresponding to a false discovery rate (FDR)<0.2 using the Benjamini–Hochberg (BH) approach as suggestive findings.

Data Availability Statement

The data generated in this study are available upon request from the corresponding author.

Results

The distribution of environmental factors in the study sample is summarized in Table 1. Compared with the control sample, breast cancer cases had a relatively higher lifetime alcohol consumption (6.5 vs. 5.7 g/day), and were less likely to be parous (85.9% vs. 88.3%). For the parous women, cases were less likely to have ever breastfed (77.5% vs. 78.7%) and reported shorter duration of

TABLE 2 Suggestive interactions between genes and environmental risk factors, with FDR-corrected adaptive weighted $P < 0.20$.

				P values			
Environmental risk factors	**Gene name**	**CHR**	**# of SNPs**	**Fixed effect**	**Random effect**	**Adaptive weighted**	**FDR-corrected, Adaptive weighted[a]**
Age at first full-term pregnancy	*C13orf45*	13q22.2	580	6.24E-01	1.03E-06	4.44E-06	0.02
Age at menarche	*RP11–219D15.3*	3q23	424	5.51E-06	1.07E-01	1.60E-05	0.08
Use of OC	*EML4*	2p21	522	4.44E-02	9.04E-05	2.91E-05	0.14
Ever breastfed	*AC114730.3*	2q37.3	192	3.33E-04	1.21E-01	6.85E-05	0.17
Ever breastfed	*AKAP3*	12p13.32	695	5.91E-04	2.51E-03	3.58E-05	0.17
Smoking status	*PMS2P3*	7q11.23	217	1.06E-05	2.80E-01	4.00E-05	0.17
Smoking status	*RP11–7I15.4*	11q14.1	350	5.94E-03	5.11E-04	6.94E-05	0.17

[a]FDR correction was conducted using the Benjamini–Hochberg (BH) approach, for each environmental factor.

breastfeeding (8.3 vs. 9.0 months). Among postmenopausal women, cases were more likely than controls to be current users of E+P MHT (18.4% vs. 12.6%) but less likely to be current users of E-only MHT (15.6% vs. 16.9%). No substantial difference was found between cases and controls for other risk factors, including age at menarche, age at first full-term birth, pre- and postmenopausal BMI, adult height, number of full-term births, OC use, and smoking status. Associations between environmental factors and breast cancer risk quantified by logistic regression are shown in the Supplementary Table S2.

The full list of GxE interaction results is reported in Supplementary Table S3.1–S3.14. Quantile–quantile plots of aMiSTi *P* values for GxE interactions are shown in Supplementary Fig. S1. We observed an inflation of interaction test statistics for current use of E-only and E+P MHT and thus, any results for MHT use should be interpreted with caution. Overall, no interactions remained statistically significant after adjusting for number of tests performed using Bonferroni correction. The strongest evidence of interaction was observed for the *C13orf45* gene on chromosome 13 and age at the first full-term pregnancy (Table 2, $P_{GXE} = 4.44 \times 10^{-6}$). The heritability of *C13orf45* expression was estimated to 0.21, based on 580 SNPs. However, the interaction was mainly driven by the random effects component ($P = 1.03 \times 10^{-6}$) rather than fixed effects component ($P = 0.62$), which indicates there may be some SNP interaction effects that are beyond the predicted gene expression. Six additional GxE interactions were identified with an FDR-corrected $P_{GXE} < 0.2$ (Table 2). These included interactions between *RP11–219D15.3* (3q23) and age at menarche ($P_{GXE} = 1.60 \times 10^{-5}$); *EML4* (2p21) and use of OCs ($P_{GXE} = 2.91 \times 10^{-5}$); history of breastfeeding and *AC114730.3* (2q37.3, $P_{GXE} = 6.85 \times 10^{-5}$) and *AKAP3* (12p13.32, $P_{GXE} = 3.58 \times 10^{-5}$) in parous women; smoking status and *PMS2P3* (7q11.23, $P_{GXE} = 4.00 \times 10^{-5}$), and *RP11–7I15.4* (11q14.1, $P_{GXE} = 6.94 \times 10^{-5}$).

Discussion

In this large transcriptome-informed investigation of GxE interactions in breast cancer, we systematically studied the interactions between predicted gene expression and fourteen behavioral and environmental risk factors. No interaction remained statistically significant after adjusting for number of tests. However, we identified seven interactions between genes and environmental factors, including age at first full-term pregnancy, age at menarche, breast feeding history, smoking status, and use of OCs, as suggestive findings with FDR-corrected $P < 0.20$. Our findings did not support a significant role played by gene expression in modifying the associations between established risk factors and breast cancer risk.

The strongest interaction identified was between the *C13orf45* gene and age at the first full-term pregnancy. *C13orf45*, or *LMO7DN*, is a long noncoding RNA (lncRNA) located downstream of the LIM domain only protein 7 (*LMO7*). Few studies have directly focused on the function of *C13orf45* gene. The expression of *LMO7* has been found to play an important role in skeletal muscle transcription and cardiac development (72–74). Irregular expression of the *LMO7* gene has been linked to multiple types of cancer, including breast, thyroid and lung (75–78). Specifically, Hu and colleagues reported that the knockdown of *LMO7* gene in the breast cancer cell line MDA-MB-231 could impair cell migration (76). In the same study, the upregulation of *LMO7* was also found in the stroma of invasive breast carcinoma, which presumably correlated with the expression of serum response factors that regulate muscle and actin cytoskeleton functions. Epidemiologic studies have consistently shown the positive association between later age at first birth and higher incidence of breast cancer (79–81), which can at least be partially explained by pregnancy-induced changes in sex hormones. Earlier differentiation of mammary epithelium induced by estrogen and progestogen at pregnancy can reduce the susceptibility of neoplastic transformation and lower the subsequent disease risk (82). However, there is no direct evidence that this mechanism might interplay with the expression of *C13orf45* or *LMO7*, and therefore, functional follow-up would be needed to explore this potential finding further.

Some of the six additional genes with an FDR-corrected $P_{interaction} < 0.2$ identified in our study have previously been linked to breast cancer development. The translocation and fusion of echinoderm microtubule-associated protein-like 4 (*EML4*) and anaplastic lymphoma kinase (*ALK*) have been implicated in various cancers. For example, the *EML4-ALK1* fusion has been observed in patients with non–small cell lung cancer (83–85), as well as in tumor samples from patients with breast and colorectal cancer (86). *ALK* gene was observed to amplify in most inflammatory breast cancer (IBC; ref. 87), a rare form of disease characterized by an early average age of diagnosis, aggressive histopathologic features, and poor survival (88). There is evidence that IBC cases has a higher prevalence of OC use than other breast cancer cases (89), which suggests that *EML4* may interact with the effect of OC use through inflammatory-related pathways. *AKAP3* is a member of A-kinase anchoring proteins, which has been recognized as a cancer-testis antigen for multiple types of cancer, including ovarian, hepatocellular, and colorectal (90–92). In an investigation of 162

tumor and normal tissues of breast, lack of *AKAP3* expression was observed to be significantly associated with triple-negative breast cancer, breast tumor size, tumor stage, and 5-year disease-free survival (93). The *PMS2P3* gene has been suggested to interact through gene expression with *PMS2* (94), a gene linked to poor survival from breast cancer (95). Noticeably, *PMS2P3* gene belongs to the mismatch repair (MMR) system, which has been observed to have a stronger effect among smokers in affecting colorectal cancer risk, relative to the never smokers (96). Further studies are needed to confirm these suggestive interactions and corresponding biological mechanisms with more direct evidence.

None of the suggestive interaction identified in our study has been observed by previous GxE studies of breast cancer. Otherwise, we were not able to replicate any significant interactions reported by the other studies, including for the genes harboring the variants with significant GxE interaction. This inconsistency could potentially be attributed to various reasons, such as different study populations, analysis approaches and importantly, adjustment for multiple testing. Given the huge number of tests (4,864 genes × 14 environmental risk factors) performed in our analysis, we performed a conservative Bonferroni correction approach and defined a threshold of $P < 7.34 \times 10^{-7}$ as statistically significant. As this stringent threshold may yield false negative results, we further adopted a more liberal threshold and reported all GxE interactions with FDR-corrected $P < 0.20$ for each environmental factor.

Our study has several strengths. First, to our knowledge this is the first study to incorporate breast tissue specific gene expression models to inform our GxE interaction analysis. Previous research has suggested that breast cancer susceptibility loci are enriched in regulatory regions identified in breast tissue or cell lines (67, 97). Based on this tissue specificity, we utilized genotype and gene expression data from mammary tissue to build gene expression prediction models, and used these models as prior information when assessing GxE interactions. By using a mixed-effects score test which enables the consideration of both fixed and random effects of the interaction, we were able to take into account the effect of genetic variants not involved in gene expression regulation. To avoid potential selection bias, we limited our study population to breast cancer cases and controls from population-based studies. However, our study was based on European ancestry women only, and thus our study conclusions may not be applicable to women with other ancestry. For certain suggestive GxE interactions detected, the results were mainly driven by the random effect component rather than the fixed effect, which made it challenging to explain the mechanisms or pathway underneath. A proportion of the studies included in our analysis adopted the case–control study design, which collected risk factor data based on self-report approaches. Consequently, the risk factor data, although centrally harmonized across all studies, might still be susceptible to recall bias. Our study did not stratify the breast cancer cases by menopausal or estrogen receptor (ER) status and investigate the subtype-specific GxE interaction, which may be a missed opportunity as the disease etiology differs across these subtypes. The results for current use of estrogen-only and estrogen plus progestogen MHT showed evidence of inflated type I error rates, indicating potential issues with distribution or modeling of those risk factors.

In conclusion, our study incorporated information on gene expression to investigate comprehensively the interactions between environmental risk factors and genetic variants on breast cancer risk using a mixed-effects score test approach. Our findings suggest a lack of evidence to demonstrate the role played by gene expression in modifying the associations between established risk factors and breast cancer risk.

Authors' Disclosures

X. Wang reports currently employed by Flatiron Health, Inc., and reports being a stockholder in Roche. Both of these relationships were after the completion of the submitted work, and were not in association with the submitted work. Y. Su reports grants from NIH during the conduct of the study. M.K. Bolla reports grants from CRUK during the conduct of the study. P.D. Pharoah reports grants from Cancer Research UK during the conduct of the study. R.A. Murphy reports personal fees from Pharmavite LLC outside the submitted work. G.G. Giles reports grants from National Health and Medical Research Council (Australia) paid to my institution Cancer Council Victoria during the conduct of the study. J. Simard reports grants from Genome Canada/Canadian Institutes of Health Research/Genome Quebec during the conduct of the study. C. Peng reports grants from NCI and grants from NIH during the conduct of the study. R.M. Tamimi reports grants from NIH/NCI during the conduct of the study. A. Eliassen reports grants from NIH during the conduct of the study. A. Wolk reports grants from Swedish Research Council and grants from Swedish Cancer Foundation during the conduct of the study. A.J. Swerdlow reports grants from Breast Cancer Now during the conduct of the study. D.F. Easton reports grants from European Union, grants from Cancer Research UK, and grants from Genome Canada during the conduct of the study. R.L. Milne reports grants from NHMRC during the conduct of the study. P. Kraft reports grants from NIH during the conduct of the study. S. Lindström reports grants from NIH during the conduct of the study. No other disclosures were reported.

Authors' Contributions

X. Wang: Conceptualization, formal analysis, investigation, methodology, writing-review and editing. **H. Chen:** Conceptualization, formal analysis, methodology, writing-original draft, writing-review and editing. **P.M. Kapoor:** Data curation, writing-review and editing. **Y.-R. Su:** Software, methodology. **M.K. Bolla:** Data curation, project administration. **J. Dennis:** Data curation, writing-review and editing. **A.M. Dunning:** Data curation. **M. Lush:** Data curation. **Q. Wang:** Data curation. **K. Michailidou:** Data curation, methodology, writing-review and editing. **P.D.P. Pharoah:** Methodology, writing-review and editing. **J.L. Hopper:** Data curation, writing-review and editing. **M.C. Southey:** Data curation. **S. Koutros:** Data curation. **L.E.B. Freeman:** Data curation. **J. Stone:** Data curation, writing-review and editing. **G. Rennert:** Data curation. **R. Shibli:** Data curation. **R.A. Murphy:** Data curation, writing-review and editing. **K. Aronson:** Data curation. **P. Guénel:** Data curation. **T. Truong:** Data curation. **L.R. Teras:** Data curation. **J.M. Hodge:** Data curation. **F. Canzian:** Data curation. **R. Kaaks:** Data curation. **H. Brenner:** Data curation. **V. Arndt:** Data curation. **R. Hoppe:** Data curation. **W.-Y. Lo:** Data curation. **, S. Behrens:** Data curation. **A. Mannermaa:** Data curation. **V.-M. Kosma:** Data curation. **A. Jung:** Data curation. **H. Becher:** Data curation. **G.G. Giles:** Data curation. **C.A. Haiman:** Data curation. **G. Maskarinec:** Data curation. **C. Scott:** Data curation. **S. Winham:** Data curation. **J. Simard:** Data curation. **M.S. Goldberg:** Data curation. **W. Zheng:** Data curation, writing-review and editing. **J. Long:** Data curation. **M.A. Troester:** Data curation. **M.I. Love:** Data curation. **C. Peng:** Data curation, writing-review and editing. **R. Tamimi:** Data curation, writing-review and editing. **H. Eliassen:** Data curation. **M. García-Closas:** Data curation. **J. Figueroa:** Data curation. **T. Ahearn:** Data curation. **R. Yang:** Data curation. **D.G. Evans:** Data curation, writing-review and editing. **A. Howell:** Data curation. **P. Hall:** Data curation, writing-review and editing. **K. Czene:** Data curation. **A. Wolk:** Data curation. **D.P. Sandler:** Data curation. **J.A.**

Taylor: Data curation. **A.J. Swerdlow:** Data curation, writing-review and editing. **N. Orr:** Data curation. **J.V. Lacey:** Data curation. **S. Wang:** Data curation. **H. Olsson:** Data curation. **D.F. Easton:** Data curation, writing-review and editing. **R.L. Milne:** Data curation, writing-review and editing. **L. Hsu:** Software, methodology, writing-review and editing. **P. Kraft:** Data curation, writing-review and editing. **J. Chang-Claude:** Resources, data curation, supervision, methodology, writing-review and editing. **S. Lindström:** Conceptualization, resources, supervision, methodology, writing-review and editing.

Acknowledgments

This work was supported by NIH R01 CA244670. We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. J.L. Hopper is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C. Southey is a NHMRC Senior Research Fellow. Full individual study acknowledgements can be found in the Supplementary Information.

Note

Supplementary data for this article are available at Cancer Research Communications Online (https://aacrjournals.org/cancerrescommun/).

Received November 05, 2021; revised March 21, 2022; accepted March 24, 2022; published first April 08, 2022.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71: 209-49.
2. Möller S, Mucci LA, Harris JR, Scheike T, Holst K, Halekoh U, et al. The heritability of breast cancer among women in the nordic twin study of cancer. Cancer Epidemiol Biomarkers Prev 2016;25: 145-50.
3. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial risk and heritability of cancer among twins in nordic countries. JAMA 2016;315: 68-76.
4. Kapoor PM, Lindström S, Behrens S, Wang X, Michailidou K, Bolla MK, et al. Assessment of interactions between 205 breast cancer susceptibility loci and 13 established risk factors in relation to breast cancer risk, in the Breast Cancer Association Consortium. Int J Epidemiol 2020;49: 216-32.
5. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol 2012;13: 1141-51.
6. Premenopausal Breast Cancer Collaborative Group, Schoemaker MJ, Nichols HB, Wright LB, Brook MN, Jones ME, et al. Association of body mass index and age with subsequent breast cancer risk in premenopausal women. JAMA Oncol 2018;4: e181771.
7. Neuhouser ML, Aragaki AK, Prentice RL, Manson JE, Chlebowski R, Carty CL, et al. Overweight, obesity, and postmenopausal invasive breast cancer risk: a secondary analysis of the women's health initiative randomized clinical trials. JAMA Oncol 2015;1: 611-21.
8. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. Lancet 2014;384: 755-65.
9. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast Cancer Res 2006;8: R43.
10. Shah NR, Borenstein J, Dubois RW. Postmenopausal hormone therapy and breast cancer: a systematic review and meta-analysis. Menopause 2005;12: 668-78.
11. Lee S, Kolonel L, Wilkens L, Wan P, Henderson B, Pike M. Postmenopausal hormone therapy and breast cancer risk: the multiethnic cohort. Int J Cancer 2006;118: 1285-91.
12. Nickels S, Truong T, Hein R, Stevens K, Buck K, Behrens S, et al. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. PLoS Genet 2013;9: e1003284.
13. Schoeps A, Rudolph A, Seibold P, Dunning AM, Milne RL, Bojesen SE, et al. Identification of new genetic susceptibility loci for breast cancer through consideration of gene-environment interactions. Genet Epidemiol 2014;38: 84-93.
14. Barrdahl M, Canzian F, Joshi AD, Travis RC, Chang-Claude J, Auer PL, et al. Post-GWAS gene-environment interplay in breast cancer: results from the breast and prostate cancer cohort consortium and a meta-analysis on 79,000 women. Hum Mol Genet 2014;23: 5260-70.
15. Campa D, Kaaks R, Le Marchand L, Haiman CA, Travis RC, Berg CD, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. J Natl Cancer Inst 2011;103: 1252-63.
16. Hein R, Flesch-Janys D, Dahmen N, Beckmann L, Lindström S, Schoof N, et al. A genome-wide association study to identify genetic susceptibility loci that modify ductal and lobular postmenopausal breast cancer risk associated with menopausal hormone therapy use: a two-stage design with replication. Breast Cancer Res Treat 2013;138: 529-42.
17. Milne RL, Gaudet MM, Spurdle AB, Fasching PA, Couch FJ, Benítez J, et al. Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: a combined case-control study. Breast Cancer Res 2010;12: R110.
18. Rudolph A, Hein R, Lindström S, Beckmann L, Behrens S, Liu J, et al. Genetic modifiers of menopausal hormone replacement therapy and breast cancer risk: a genome-wide interaction study. Endocr Relat Cancer 2013;20: 875-87.
19. Rudolph A, Milne RL, Truong T, Knight JA, Seibold P, Flesch-Janys D, et al. Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors. Int J Cancer 2015;136: E685-96.
20. Travis RC, Reeves GK, Green J, Bull D, Tipper SJ, Baker K, et al. Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the million women study. Lancet 2010;375: 2143-51.
21. Rudolph A, Chang-Claude J, Schmidt MK. Gene-environment interaction and risk of breast cancer. Br J Cancer 2016;114: 125-33.
22. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet 2015;16: 197-212.
23. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet 2014;95: 535-52.
24. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012;337: 1190-5.
25. Wu L, Shi W, Long J, Guo X, Michailidou K, Beesley J, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet 2018;50: 968-78.
26. Su YR, Di CZ, Hsu L, Genetics and Epidemiology of Colorectal Cancer Consortium. A unified powerful set-based test for sequencing data analysis of GxE interactions. Biostatistics 2017;18: 119-31.
27. Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MRE, et al. Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. J Natl Cancer Inst 2003;95: 448-57.

28. Koutros S, Alavanja MCR, Lubin JH, Sandler DP, Hoppin JA, Lynch CF, et al. An update of cancer incidence in the agricultural health study. J Occup Environ Med 2010;52: 1098-105.
29. Fritschi L, Erren TC, Glass DC, Girschik J, Thomson AK, Saunders C, et al. The association between different night shiftwork factors and breast cancer: a case-control study. Br J Cancer 2013;109: 2472-80.
30. Rennert G, Lejbkowicz F, Cohen I, Pinchev M, Rennert HS, Barnett-Griness O. MutYH mutation carriers have increased breast cancer risk. Cancer 2012;118: 1989-93.
31. Rennert G, Pinchev M, Rennert HS. Use of bisphosphonates and risk of postmenopausal breast cancer. J Clin Oncol 2010;28: 3577-81.
32. Kobayashi LC, Janssen I, Richardson H, Lai AS, Spinelli JJ, Aronson KJ. A case-control study of lifetime light intensity physical activity and breast cancer risk. Cancer Causes Control 2014;25: 133-40.
33. Kobayashi LC, Janssen I, Richardson H, Lai AS, Spinelli JJ, Aronson KJ. Moderate-to-vigorous intensity physical activity across the life course and risk of pre- and post-menopausal breast cancer. Breast Cancer Res Treat 2013;139: 851-61.
34. Grundy A, Richardson H, Burstyn I, Lohrisch C, SenGupta SK, Lai AS, et al. Increased risk of breast cancer associated with long-term shift work in Canada. Occup Environ Med 2013;70: 831-8.
35. Grundy A, Schuetz JM, Lai AS, Janoo-Gilani R, Leach S, Burstyn I, et al. Shift work, circadian gene variants and risk of breast cancer. Cancer Epidemiol 2013;37: 606-12.
36. Menegaux F, Truong T, Anger A, Cordina-Duverger E, Lamkarkach F, Arveux P, et al. Night work and breast cancer: a population-based case-control study in France (the CECILE study). Int J Cancer 2013;132: 924-31.
37. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, et al. The American cancer society cancer prevention study II nutrition cohort: rationale, study design, and baseline characteristics. Cancer 2002;94: 2490-501.
38. Bernstein L, Allen M, Anton-Culver H, Deapen D, Horn-Ross PL, Peel D, 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.
39. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;5: 1113-24.
40. Widschwendter M, Apostolidou S, Raum E, Rothenbacher D, Fiegl H, Menon U, et al. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. PLoS One 2008;3: e2656.
41. Pesch B, Ko Y, Brauch H, Hamann U, Harth V, Rabstein S, et al. Factors modifying the association between hormone-replacement therapy and breast cancer risk. Eur J Epidemiol 2005;20: 699-711.
42. Justenhoven C, Pierl CB, Haas S, Fischer HP, Baisch C, Hamann U, et al. The CYP1B1_1358_GG genotype is associated with estrogen receptor-negative breast cancer. Breast Cancer Res Treat 2008;111: 171-7.
43. Chang-Claude J, Eby N, Kiechle M, Bastert G, Becher H. Breastfeeding and breast cancer risk by age 50 among women in Germany. Cancer Causes Control 2000;11: 687-95.
44. Gabrielson M, Eriksson M, Hammarström M, Borgquist S, Leifland K, Czene K, et al. Cohort profile: the karolinska mammography project for risk prediction of breast cancer (KARMA). Int J Epidemiol 2017;46: 1740-1g.
45. Hartikainen JM, Tuhkanen H, Kataja V, Eskelinen M, Uusitupa M, Kosma VM, et al. Refinement of the 22q12-q13 breast cancer–associated region: evidence of TMPRSS6 as a candidate gene in an eastern Finnish population. Clin Cancer Res 2006;12: 1454-62.
46. Hartikainen JM, Tuhkanen H, Kataja V, Dunning AM, Antoniou A, Smith P, et al. An autosome-wide scan for linkage disequilibrium-based association in sporadic breast cancer cases in eastern Finland: three candidate regions found. Cancer Epidemiol Biomarkers Prev 2005;14: 75-80.
47. Flesch-Janys D, Slanger T, Mutschelknauss E, Kropp S, Obi N, Vettorazzi E, et al. Risk of different histological types of postmenopausal breast cancer by type and regimen of menopausal hormone therapy. Int J Cancer 2008;123: 933-41.
48. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK, et al. Cohort profile: the melbourne collaborative cohort study (Health 2020). Int J Epidemiol 2017;46: 1757-i.
49. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151: 346-57.
50. Olsson HL, Ingvar C, Bladström A. Hormone replacement therapy containing progestins and given continuously increases breast carcinoma risk in Sweden. Cancer 2003;97: 1387-92.
51. Nielsen K, Måsbäck A, Olsson H, Ingvar C. A prospective, population-based study of 40,000 women regarding host factors, UV exposure and sunbed use in relation to risk and anatomic site of cutaneous melanoma. Int J Cancer 2012;131: 706-15.
52. Olson JE, Sellers TA, Scott CG, Schueler BA, Brandt KR, Serie DJ, 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.
53. Barrdahl M, Rudolph A, Hopper JL, Southey MC, Broeks A, Fasching PA, et al. Gene-environment interactions involving functional variants: results from the breast cancer association consortium. Int J Cancer 2017;141: 1830-40.
54. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009;41: 324-8.
55. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, et al. The carolina breast cancer study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 1995;35: 51-60.
56. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, et al. HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast Cancer Res Treat 2003;79: 355-64.
57. Hankinson SE, Willett WC, Manson JE, Colditz GA, Hunter DJ, Spiegelman D, et al. Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. J Natl Cancer Inst 1998;90: 1292-9.
58. Tworoger SS, Missmer SA, Eliassen AH, Spiegelman D, Folkerd E, Dowsett M, et al. The association of plasma DHEA and DHEA sulfate with breast cancer risk in predominantly premenopausal women. Cancer Epidemiol Biomarkers Prev 2006;15: 967-71.
59. García-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, et al. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. Hum Genet 2006;119: 376-88.
60. Pfeiffer RM, Park Y, Kreimer AR, Lacey JV Jr, Pee D, Greenlee RT, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. PLoS Med 2013;10: e1001492.
61. Evans DG, Astley S, Stavrinos P, Harkness E, Donnelly LS, Dawe S, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and family history clinics: a dual cohort study, Programme Grants for Applied Research. Southampton (UK) 2016.
62. Wedrén S, Lovmar L, Humphreys K, Magnusson C, Melhus H, Syvänen AC, et al. Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study. Breast Cancer Res 2004;6: R437-49.
63. Xu Z, Bolick SC, DeRoo LA, Weinberg CR, Sandler DP, Taylor JA. Epigenome-wide association study of breast cancer using prospectively collected sister study samples. J Natl Cancer Inst 2013;105: 694-700.
64. Nichols HB, Baird DD, DeRoo LA, Kissling GE, Sandler DP. Tubal ligation in relation to menopausal symptoms and breast cancer risk. Br J Cancer 2013;109: 1291-5.
65. Suzuki R, Ye W, Rylander-Rudqvist T, Saji S, Colditz GA, Wolk A. Alcohol and postmenopausal breast cancer risk defined by estrogen and progesterone receptor status: a prospective cohort study. J Natl Cancer Inst 2005;97: 1601-8.
66. Swerdlow AJ, Jones ME, Schoemaker MJ, Hemming J, Thomas D, Williamson J, et al. The breakthrough generations study: design of a long-term UK cohort study to investigate breast cancer aetiology. Br J Cancer 2011;105: 911-7.

67. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551: 92-4.
68. Consortium G. The genotype-tissue expression (GTEx) project. Nat Genet 2013;45: 580-5.
69. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88: 76-82.
70. Amadou A, Hainaut P, Romieu I. Role of obesity in the risk of breast cancer: lessons from anthropometry. J Oncol 2013;2013: 906495.
71. Su YR, Di C, Bien S, Huang L, Dong X, Abecasis G, et al. A mixed-effects model for powerful association tests in integrative functional genomics. Am J Hum Genet 2018;102: 904-19.
72. Dedeic Z, Cetera M, Cohen TV, Holaska JM. Emerin inhibits Lmo7 binding to the Pax3 and MyoD promoters and expression of myoblast proliferation genes. J Cell Sci 2011;124: 1691-702.
73. Mewborn SK, Puckelwartz MJ, Abuisneineh F, Fahrenbach JP, Zhang Y, MacLeod H, et al. Altered chromosomal positioning, compaction, and gene expression with a lamin A/C gene mutation. PLoS One 2010;5: e14342.
74. Mull A, Kim G, Holaska JM. LMO7-null mice exhibit phenotypes consistent with emery-dreifuss muscular dystrophy. Muscle Nerve 2015;51: 222-8.
75. He H, Li W, Yan P, Bundschuh R, Killian JA, Labanowska J, et al. Identification of a recurrent LMO7-BRAF fusion in papillary thyroid carcinoma. Thyroid 2018;28: 748-54.
76. Hu Q, Guo C, Li Y, Aronow BJ, Zhang J. LMO7 mediates cell-specific activation of the Rho-myocardin-related transcription factor-serum response factor pathway and plays an important role in breast cancer cell migration. Mol Cell Biol 2011;31: 3223-40.
77. Karlsson T, Kvarnbrink S, Holmlund C, Botling J, Micke P, Henriksson R, et al. LMO7 and LIMCH1 interact with LRIG proteins in lung cancer, with prognostic implications for early-stage disease. Lung Cancer 2018;125: 174-84.
78. Nakamura H, Hori K, Tanaka-Okamoto M, Higashiyama M, Itoh Y, Inoue M, et al. Decreased expression of LMO7 and its clinicopathological significance in human lung adenocarcinoma. Exp Ther Med 2011;2: 1053-7.
79. Bernstein L. Epidemiology of endocrine-related risk factors for breast cancer. J Mammary Gland Biol Neoplasia 2002;7: 3-15.
80. Lord SJ, Bernstein L, Johnson KA, Malone KE, McDonald JA, Marchbanks PA, et al. Breast cancer risk and hormone receptor status in older women by parity, age of first birth, and breastfeeding: a case-control study. Cancer Epidemiol Biomarkers Prev 2008;17: 1723-30.
81. Newcomb PA, Trentham-Dietz A, Hampton JM, Egan KM, Titus-Ernstoff L, Warren Andersen S, et alet al. Late age at first full term birth is strongly associated with lobular breast cancer. Cancer 2011;117: 1946-56.
82. Russo IH, Russo J. Pregnancy-induced changes in breast cancer risk. J Mammary Gland Biol Neoplasia 2011;16: 221-33.
83. Choi YL, Takeuchi K, Soda M, Inamura K, Togashi Y, Hatano S, et al. Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. Cancer Res 2008;68: 4971-6.
84. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448: 561-6.
85. Tan S, Sun D, Pu W, Gou Q, Guo C, Gong Y, et al. Circular RNA F-circEA-2a derived from EML4-ALK fusion gene promotes cell migration and invasion in non-small cell lung cancer. Mol Cancer 2018;17: 138.
86. Lin E, Li L, Guan Y, Soriano R, Rivers CS, Mohan S, et al. Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. Mol Cancer Res 2009;7: 1466-76.
87. Tuma RS. ALK gene amplified in most inflammatory breast cancers. J Natl Cancer Inst 2012;104: 87-8.
88. Dawood S, Merajver SD, Viens P, Vermeulen PB, Swain SM, Buchholz TA, et al. International expert panel on inflammatory breast cancer: consensus statement for standardized diagnosis and treatment. Ann Oncol 2011;22: 515-23.
89. Moslehi R, Freedman E, Zeinomar N, Veneroso C, Levine PH. Importance of hereditary and selected environmental risk factors in the etiology of inflammatory breast cancer: a case-comparison study. BMC Cancer 2016;16: 334.
90. Hasegawa K, Ono T, Matsushita H, Shimono M, Noguchi Y, Mizutani Y, et al. A-kinase anchoring protein 3 messenger RNA expression in ovarian cancer and its implication on prognosis. Int J Cancer 2004;108: 86-90.
91. Sharma S, Qian F, Keitz B, Driscoll D, Scanlan MJ, Skipper J, et al. A-kinase anchoring protein 3 messenger RNA expression correlates with poor prognosis in epithelial ovarian cancer. Gynecol Oncol 2005;99: 183-8.
92. Song MH, Choi KU, Shin DH, Lee CH, Lee SY. Identification of the cancer/testis antigens AKAP3 and CTp11 by SEREX in hepatocellular carcinoma. Oncol Rep 2012;28: 1792-8.
93. Esmaeili R, Majidzadeh AK, Farahmand L, Ghasemi M, Salehi M, Khoshdel AR. AKAP3 correlates with triple negative status and disease free survival in breast cancer. BMC Cancer 2015;15: 681.
94. Li X, Scanlon MJ, Yu J. Evolutionary patterns of DNA base composition and correlation to polymorphisms in DNA repair systems. Nucleic Acids Res 2015;43: 3614-25.
95. Anurag M, Punturi N, Hoog J, Bainbridge MN, Ellis MJ, Haricharan S. Comprehensive profiling of DNA repair defects in breast cancer identifies a novel class of endocrine therapy resistance drivers. Clin Cancer Res 2018;24: 4887-99.
96. Yang P, Cunningham JM, Halling KC, Lesnick TG, Burgart LJ, Wiegert EM, et al. Higher risk of mismatch repair-deficient colorectal cancer in alpha(1)-antitrypsin deficiency carriers and cigarette smokers. Mol Genet Metab 2000;71: 639-45.
97. Chen H, Kichaev G, Bien SA, MacDonald JW, Wang L, Bammler TK, et al. Genetic associations of breast and prostate cancer are enriched for regulatory elements identified in disease-related tissues. Hum Genet 2019;138: 1091-104.