Title
Transcriptome-wide association study of breast cancer risk by estrogen-receptor status

Permalink
https://escholarship.org/uc/item/6hm5f5b2

Authors
Feng, Helian
Gusev, Alexander
Pasaniuc, Bogdan
et al.

Publication Date
2020

DOI
10.1002/gepi.22288

Peer reviewed
Transcriptome-wide association study of breast cancer risk by estrogen-receptor status

Helian Feng1,2,3 | Alexander Gusev4 | Bogdan Pasaniuc5 | Lang Wu6
Jirong Long7 | Zomoroda Abu-full8 | Kristiina Aittomäki9 | Irene L. Andrulis10,11
Hoda Anton-Culver12 | Antonis C. Antoniou13 | Adalgeir Arason14,15
Volker Arndt16 | Kristan J. Aronson17 | Banu K. Arun18 | Ella Asseryanis19
Paul L. Auer20,21 | Jacopo Azzollini22 | Judith Balmaña23
Rosa B. Barkardottir14,15 | Daniel R. Barnes13 | Daniel Barrowdale13
Matthias W. Beckmann24 | Sabine Behrens25 | Javier Benitez26,27
Marina Bermisheva28 | Katarzyna Bialkowska29 | Ana Blanco26,30,31
Carl Blomqvist32,33 | Bram Boeckx34,35 | Natalia V. Bogdanova36,37,38
Stig E. Bojesen39,40,41 | Manjeet K. Bolla13 | Bernardo Bonanni42 | Ake Borg43
Hiltrud Brauch44,45,46 | Hermann Brenner46,47 | Ignacio Briceno48,49
Annegien Broeck30 | Thomas Brüning51 | Barbara Burwinkel52,53 | Qiuyin Cai7
Trinidad Caldés54 | Maria A. Caligo55 | Ian Campbell56,57 | Sander Canisius50,58
Daniele Campa59 | Brian D. Carter60 | Jonathan Carter61 | Jose E. Castelao62
Jenny Chang-Claude25,63 | Stephen J. Chanock64 | Hans Christiansen66
Wendy K. Chung65 | Kathleen B. M. Claes66 | Christine L. Clarke67
GEMO Study Collaborators68,69,70 | EMBRACE Collaborators13
GC-HBOC study Collaborators71 | Fergus J. Couch72 | Angela Cox73
Simon S. Cross74 | Cezary Cybulski20 | Kamila Czene75 | Mary B. Daly76
Miguel de la Hoya54 | Kim De Leeneer66 | Joe Dennis13 | Peter Devilee77,78
Orland Diez79,80 | Susan M. Domchek81 | Thilo Dörk37 | Isabel dos-Santos-Silva82
Alison M. Dunning83 | Miriam Dwek84 | Diana M. Eccles85 | Bent Ejlersen86
Carolina Ellberg87 | Christoph Engel88,89 | Mikael Eriksson75
Peter A. Fasching24,90 | Olivia Fletcher91 | Henrik Flyger92 | Florentia Fostira93
Eitan Friedman94,95 | Lin Fritschi96 | Debra Frost13 | Marieke Gabrielson75
Patricia A. Ganz97 | Susan M. Gapstur60 | Judy Garber98
Montserrat García-Closas64,99,100 | José A. García-Sáenz54 | Mia M. Gaudet60
Graham G. Giles101,102 | Gord Glendon10 | Andrew K. Godwin103
Mark S. Goldberg104,105 | David E. Goldgar106 | Anna González-Neira27

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. Genetic Epidemiology published by Wiley Periodicals, Inc.
Ontario Cancer Genetics Network, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada
Lombardi Comprehensive Cancer Center, Georgetown University, Washington, District of Columbia
Research Centre for Genetic Engineering and Biotechnology 'Georgi D, Efremov', Macedonian Academy of Sciences and Arts, Skopje Republic of North Macedonia, North Macedonia
Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland
Parkville Familial Cancer Centre, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia
State Research Institute Innovative Medicine Center, Vilnius, Lithuania
Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
Clinical Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York
David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California, Los Angeles, California
Department of Genetics and Fundamental Medicine, Bashkir State Medical University, Ufa, Russia
Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway
Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
Molecular Diagnostic Unit, Hereditary Cancer Program, ICO-IDIBELL (Bellvitge Biomedical Research Institute, Catalan Institute of Oncology), CIBERONC, Barcelona, Spain
Departement de Gene netique, CHU de Grenoble, Grenoble, France
Genetic Epidemiology of Cancer Team, Inserm U900, Paris, France
Department of Health Sciences Research, Mayo Clinic, Scottsdale, Arizona
Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington
Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington
Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland
Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland
Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland
Department of Clinical Science and Education, Sodersjukhuset, Karolinska Institutet, Stockholm, Sweden
Moore Cancer Center, University of California San Diego, La Jolla, California
Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California
Immunology and Molecular Oncology Unit, Veneto Institute of Oncology EIOV—IRCCS, Padua, Italy
Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece
Department of Gynecology and Obstetrics, Ludwig Maximilian University of Munich, Munich, Germany
MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, University College London, London, UK
Department of Electron Microscopy/Molecular Pathology and The Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus
Faculty of Medicine, University of Heidelberg, Heidelberg, Germany
NRG Oncology, Statistics and Data Management Center, Roswell Park Cancer Institute, Buffalo, New York
Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada
Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada
Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, California
Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium
Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
Latvian Biomedical Research and Study Centre, Riga, Latvia
Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York
Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary
Center for Clinical Cancer Genetics, The University of Chicago, Chicago, Illinois
Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
Section of Molecular Diagnostics, Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark
Department of Genetics, Portuguese Oncology Institute, Porto, Portugal
Genome Diagnostics Program, IFOM—The FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy
Department of Medicine Oncology, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia

FENG ET AL.
Correspondence
Xia Jiang, Harvard T. H. Chan School of Public Health, Building 2-205, 677 Huntington Avenue, Boston, MA 02115.
Email: xiajiang@hsph.harvard.edu

Abstract
Previous transcriptome-wide association studies (TWAS) have identified breast cancer risk genes by integrating data from expression quantitative loci and genome-wide association studies (GWAS), but analyses of breast cancer subtype-specific associations have been limited. In this study, we conducted a TWAS using gene expression data from GTEx and summary statistics from the hitherto largest GWAS meta-analysis conducted for breast cancer overall, and by estrogen receptor subtypes (ER+ and ER−). We further compared associations with ER+ and ER− subtypes, using a case-only TWAS approach. We also conducted multigene conditional analyses in regions with multiple TWAS associations. Two genes, STXBP4 and HIST2H2BA, were specifically associated with ER+ but not with ER− breast cancer. We further identified 30 TWAS-significant genes associated with overall breast cancer risk, including four that were not identified in previous studies. Conditional analyses identified single independent breast-cancer gene in three of six regions harboring multiple TWAS-significant genes. Our study provides new information on breast cancer genetics and biology, particularly about genomic differences between ER+ and ER− breast cancer.

KEYWORDS
breast cancer subtype, causal gene, GWAS, TWAS

1 | INTRODUCTION

Breast cancer is the most common malignancy among women worldwide (Bray et al., 2018). The disease has a strong inherited component (Beggs & Hodgson, 2009); linkage studies have identified infrequent mutations in BRCA1/2 (Easton et al., 2007; Seal et al., 2006; Turnbull et al., 2010) and genome-wide association studies (GWAS) have identified 177 susceptibility loci to date (Michailidou et al., 2017). However, these GWAS-discovered variants explain only 18% of the familial relative risk of breast cancer. Moreover, the causal mechanism driving GWAS associations remains largely unknown, as many variants are located in noncoding or intergenic regions, and are not in strong linkage disequilibrium (LD) with known protein-coding variants (Beggs & Hodgson, 2009; Michailidou et al., 2015).

Breast cancer is a heterogeneous disease consisting of several well-established subtypes. One of the most important markers of breast cancer subtypes is estrogen receptor (ER) status. ER+ and ER− tumors differ in etiology (X. R. Yang, Chang-Claude, et al., 2011), genetic predisposition (Mavaddat, Antoniou, Easton, & Garcia-Closas, 2010), and clinical behavior (Blows et al., 2010). ER− tumor occurs more often among younger women, and patients are more likely to carry BRCA1 pathogenic variants (Atchley et al., 2008; Garcia-Closas et al., 2013). ER− tumor also has worse short-term prognosis. Among the 177 GWAS-identified breast cancer-associated single nucleotide polymorphisms (SNPs), around 50 are more strongly associated with ER+ disease and 20 are more strongly associated with ER− disease (Michailidou et al., 2017; Milne et al., 2017).

SNPs associated with complex traits are more likely to be in regulatory regions than in protein-coding regions, and many of these SNPs are also associated with expression levels of nearby genes (Nicolae et al., 2010). For example, breast cancer GWAS-identified variants at 6q25.1 regulate ESR1, but also coregulate other local genes such as RMND1, ARMT1, and CCDC170 (Dunning et al., 2016, p. 1). These results suggest that by integrating genotype, phenotype, and gene expression, we can identify novel trait-associated genes and understand biological mechanisms. However, due to costs and tissue
availability, acquiring GWAS and gene expression data for the same set of individuals remains challenging.

A recently published approach, referred to as transcriptome-wide association study (TWAS; Gamazon et al., 2015; Gusev et al., 2016), overcomes these difficulties by using a relatively small set of reference individuals for whom both gene expression and SNPs have been measured to impute the cis-genetic component of expression for a much larger set of individuals from their GWAS summary statistics. The association between the predicted gene expression and traits can then be tested. This method has been shown to have greater power relative to GWAS; and has identified 1,196 trait-associated genes across 30 complex traits in a recently performed multitissue TWAS (Mancuso et al., 2017).

To date, three TWAS of breast cancer have been conducted (Gao, Pierce, Olopade, Im, & Huo, 2017; Hoffman et al., 2017; Wu et al., 2018). A fourth study linked expression quantitative loci (eQTL) data across multiple tissues and breast cancer GWAS results using EUGENE, a statistical approach that sums evidence for association with disease across eQTLs regardless of directionality. That study then tested EUGENE-significant genes using a TWAS statistic, which does take directionality into account (Ferreira et al., 2019). The two earliest TWAS used GWAS data from the National Cancer Institute’s “Up for a Challenge” competition, which included data from 12,100 breast cancer cases (of which 3,900 had ER− disease) and 11,400 controls, as well as eQTL data from breast tissue and whole blood from the GTEx and DGN projects (Gao et al., 2017; Hoffman et al., 2017). The subsequent TWAS by Wu et al. (2018) and the EUGENE analysis by Ferreira et al. (2019) used results from a much larger GWAS conducted by the Breast Cancer Association Consortium (BCAC), which included 122,977 cases (of which 21,468 had ER− disease) and 105,974 controls. Together, these four studies have identified 59 genes whose predicted expression levels are associated with risk of overall breast cancer, and five associated with risk of ER− disease. Of these 64 genes, 30 are at loci not previously identified by breast cancer GWAS.

These previous TWAS largely focused on overall breast cancer risk. Analyses of ER− disease either were conducted using a small sample size (Gao et al., 2017) or did not scan all genes using a directional TWAS approach (Ferreira et al., 2019). Moreover, none of the previous analyses considered ER+ disease specifically or examined differences in association between predicted gene expression and ER+ versus ER− disease.

The interpretation of TWAS results is not straightforward (Wainberg et al., 2019). Specifically, TWAS statistic by itself cannot distinguish between a mediated effect (SNPs influence breast cancer risk by changing the expression of the tested gene), pleiotropy (SNPs associated with gene expression also influence breast cancer risk through another mechanism), or colocalization (SNPs associated with gene expression are in LD with other SNPs that influence breast cancer risk through another mechanism). Previous studies have conducted limited sensitivity analyses (e.g., Wu et al., 2018 and Ferreira et al., 2019) conditioned the TWAS tests on lead GWAS SNPs, but the genetic architecture at TWAS-identified loci remains largely unclear.

In the current analysis, we complement previous work by conducting a TWAS for overall breast cancer and for ER+ and ER− subtypes. We also applied a case-only TWAS test to identify predicted transcript levels that were differentially associated with ER+ and ER− disease. We conducted expanded sensitivity analyses, conditioning on multiple TWAS-significant genes in a region to account for possible confounding due to LD (colocalization). We chose to focus on the expression of normal breast tissue of European ancestry women to maximize specificity and identify good targets for near-term follow-up experiments in mammary cells. One advantage of using a biologically relevant tissue is that it both increases the a priori plausibility of observed associations and increases the likelihood that genes with observed associations will be expressed and influence tumor development in cells from the target tissue. We have reproduced previous results (Ferreira et al., 2019; Wu et al., 2018) and provided evidence regarding the independent associations of multiple genes in regions containing one or more TWAS-significant genes. We also identified genes with subtype-specific associations, highlighting different biological mechanisms likely underlying the disease subtypes.

2 | MATERIAL AND METHODS

2.1 | Gene expression reference panel

The transcriptome and high-density genotyping data used to build the gene expression model (reference panel) were retrieved from GTEx (GTEx Consortium, 2015), a consortium collected high-quality gene expression RNA-seq data across 44 body sites from 449 donors, and genome-wide genetic information. For the current study, we included 67 women of European ancestry who provided normal breast mammary tissues. RNA samples extracted from tissues were sequenced to generate data on 12,696 transcripts. Genomic DNA samples were genotyped using Illumina OMNI 5 M or 2.5 M arrays, processed with a standard GTEx protocol. Briefly, SNPs with call rates <98%, with differential missingness between the two array experiments (5 M or 2.5 M), with Hardy-Weinberg equilibrium p < 10^-6 or showing batch
effects were excluded. The genotypes were then imputed to the HaploPy Reference Consortium reference panel (McCarthy et al., 2016) using Minimac3 for imputation and SHAPEIT for pre-phasing (Delaneau, Marchini, & Zagury, 2011; Howie, Donnelly, & Marchini, 2009). Only SNPs with high imputation quality \( r^2 \geq 0.8 \), minor allele frequency (MAF) \( \geq 0.05 \), and were included in the HapMap Phase 2 version were used to build the expression prediction models.

### 2.2 Breast cancer meta-GWAS data

The GWAS breast cancer summary-level data were mainly provided by the Breast Cancer Association Consortium (BCAC; Michailidou et al., 2017), as well as the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). BCAC conducted the largest breast cancer meta-GWAS to date (referred as the overall breast cancer GWAS analysis). The BCAC included 122,977 cases and 105,974 controls of European ancestry. Among these, 46,785 cases and 42,892 controls were genotyped using the Illumina iSelect genotyping array (iCOGS) on 211,155 SNPs; and 61,282 cases and 45,494 controls were genotyped using the Illumina OncoArray on 570,000 SNPs (Yovel, Franz, Stilz, & Schnitzler, 2008). The study also included data from 11 other GWAS on 14,910 cases and 17,588 controls. Genetic data for all individual participating studies were imputed to the 1000 Genomes Project Phase 3 v5 EUR reference panel. Logistic regression was fitted to estimate per-allele odds ratios (ORs), adjusting for country and top principal components (PCs). Inverse variance fixed-effect meta-analysis was used to combine the genetic association for breast cancer risk across studies (Milne et al., 2017). In CIMBA, genotypes were generated by the Illumina OncoArray and imputed to the 1000 Genomes Project Phase 3 v5 EUR reference panel (Amos et al., 2016). A retrospective cohort analysis framework was adopted to estimate per-allele hazard ratios (HRs), modelling time-to-breast-cancer and stratified by country, Ashkenazi Jewish origin and birth cohort (Antoniou et al., 2005; Barnes et al., 2012). Fixed-effect meta-analysis (Willer, Li, & Abecasis, 2010) was performed to combine results across genotyping initiatives within the two consortia, assuming that the OR and HR estimates had roughly the same underlying relative risk. We restricted subsequent analyses to SNPs with an imputation \( r^2 > 0.3 \), and a MAF > 0.005 across all platforms were included in the analysis (approximately 11.5 M).

For the ER+ subtype, meta-GWAS summary data based on 69,501 ER+ cases and 105,974 controls (part of the overall breast cancer samples) were included and analyzed (Mavaddat et al., 2015). For the ER− subtype, meta-GWAS summary data based on 21,468 ER− cases and 105,974 controls from the BCAC were combined with 9,414 additional BRCA1 mutation-positive cases and 9,494 BRCA1 mutation-positive controls from CIMBA (Milne et al., 2017).

To distinguish different genetic signals between ER+ and ER− subtypes, we further retrieved GWAS summary-level data on a case-only GWAS, which compared ER+ patients (sample size: 23,330 in iCOGs and 44,746 in OncoArray) to ER− patients (sample size: 5,479 and 11,856; Milne et al., 2017). Logistic regression was performed to test the association between genetic variants with known ER status in the two studies separately, adjusting for substudy and top PCs for iCOGs, and patients’ countries and top PCs for OncoArray. Results were then combined using a fixed-effect meta-analysis.

### 2.3 Constructing expression weights

Before constructing the expression model (using GTEx data, regress gene expression on SNPs), we set several criteria to select eligible candidate genes for inclusion in the model (from the total 12,696 transcripts). We used a REML algorithm implemented in GCTA to estimate the heritability \( h^2 \) for each transcript expression (Cai et al., 2014; J. Yang et al., 2010). Only genes with significant heritability (nominal \( p \leq 0.01 \)) were included in the subsequent model construction (J. Yang, Lee, Goddard, & Visscher, 2011). The \( p \) values for null hypotheses \( cish^2_g = 0 \) were computed using a likelihood ratio test. To account for population stratification, 20 PCs were always included as fixed effects. Consistent with previous research (ENCODE Project Consortium, 2012; J. Yang et al., 2010), we observed strong evidence for \( cish^2_g \) on many genes (significantly non-zero for 1,355 genes).

We then constructed linear genetic predictors of gene expression for these genes. We performed five models: Bayesian Sparse Linear Mixed model, Best Linear Unbiased Predictor model, Elastic-net regression (with mixing parameter of 0.5), LASSO regression and Single best eQTL model. We used a fivefold cross-validation strategy to validate each model internally. Only genes with good model performance, corresponding to a prediction \( R^2 \) value (the square of the correlation between predicted and observed expression) of at least 1% (0.10 correlation) in at least one of the five models, were included in subsequent TWAS analyses. The weights were chosen from the best performed model out of the five models. We adopted this additional filter to improve the interpretability and specificity of results: significant TWAS results based on models with little or no predictive ability likely result from pleiotropy or
| Cytoband | Gene          | Chromosome: Position (start-end) | Number of SNPs | Heritability | Cross validation $r^2$ | TWAS p values |  |
|----------|---------------|----------------------------------|----------------|--------------|------------------------|----------------|---|
| p11.2    | HIST2H2BA     | 1: 12090628–12091507             | 77             | 0.10         | 0.04                   | 3.1E–30        | 2.0E–32 | 3.6E–01 | 1.4E–07 |
| q21.1    | NUDT17        | 1: 145586804–145589439            | 110            | 0.07         | 0.03                   | 1.7E–09        | 8.8E–10 | 2.4E–01 | 3.3E–03 |
| q33.1    | ALS2CR12      | 2: 202152994–202222121           | 363            | 0.35         | 0.03                   | 2.2E–11        | 8.4E–07 | 2.5E–06 | 8.8E–01 |
|          | CASP8         | 2: 202098166–202152434           | 371            | 0.32         | 0.15                   | 1.8E–07        | 8.2E–06 | 2.9E–04 | 7.6E–01 |
| q25.31   | LINC00886     | 3: 1156465135–156534851          | 452            | 0.17         | 0.04                   | 4.7E–05        | 1.9E–04 | 3.2E–03 | 7.3E–01 |
| p16.3    | MAEA          | 4: 1283639–1333925               | 374            | 0.42         | 0.23                   | 3.9E–05        | 1.6E–04 | 2.3E–01 | 2.3E–01 |
| q14.2    | ATG10         | 5: 81267844–81551958             | 520            | 0.44         | 0.26                   | 1.9E–10        | 2.4E–10 | 2.7E–01 | 2.4E–02 |
|          | ATP6AP1L      | 5: 81575281–81682796             | 467            | 0.62         | 0.51                   | 2.3E–07        | 3.3E–06 | 2.9E–02 | 2.3E–01 |
| q14.1    | RP11–250B2.5  | 6: 81176675–81178797             | 402            | 0.12         | 0.08                   | 6.9E–07        | 7.3E–03 | 5.9E–03 | 8.2E–01 |
| q22.33   | RP11–730E3.3  | 6: 130454555–1130465515          | 557            | 0.33         | 0.11                   | 4.5E–12        | 1.2E–06 | 5.2E–09 | 5.5E–02 |
|          | L3MBTL3       | 6: 130334844–130462594           | 609            | 0.31         | 0.23                   | 8.5E–14        | 3.6E–08 | 1.8E–10 | 1.4E–01 |
| p15.1    | GDI2          | 10: 5807186–584095               | 727            | 0.29         | 0.01                   | 4.8E–07        | 3.1E–07 | 4.0E–01 | 2.1E–01 |
| p15.5    | MRPL23-AS1    | 11: 2004467–2011150              | 449            | 0.34         | 0.09                   | 7.4E–06        | 1.4E–04 | 2.9E–01 | 2.2E–01 |
| q13.2    | RP11–554A11.9 | 11: 68923378–68927220            | 428            | 0.23         | 0.08                   | 1.5E–06        | 4.9E–04 | 7.1E–03 | 6.9E–01 |
| q15      | NUP107        | 12: 69082742–69136785            | 538            | 0.27         | 0.05                   | 6.1E–06        | 8.2E–07 | 4.3E–01 | 1.0E–01 |
| q24.3    | ULK3          | 15: 75128457–75135538            | 286            | 0.19         | 0.15                   | 3.9E–05        | 2.1E–05 | 2.8E–02 | 5.2E–01 |
|          | MAN2C1        | 15: 75648133–75659986            | 226            | 0.37         | 0.34                   | 7.4E–07        | 1.7E–04 | 1.8E–03 | 6.3E–01 |
|          | CTD-232K18.1  | 15: 75819491–75893546            | 226            | 0.36         | 0.19                   | 7.5E–07        | 3.9E–04 | 2.3E–03 | 3.1E–01 |
| q21.31   | HSD17B1P1     | 17: 40698782–40700724            | 245            | 0.09         | 0.07                   | 1.4E–06        | 1.8E–03 | 6.4E–03 | 3.9E–01 |
| q21.32   | LRRC37A4P     | 17: 43578685–43623042            | 149            | 0.34         | 0.35                   | 3.1E–10        | 1.4E–06 | 2.2E–02 | 8.0E–01 |
|          | CRHR1-IT1     | 17: 43697694–43725582            | 87             | 0.35         | 0.41                   | 2.9E–10        | 1.8E–06 | 2.7E–02 | 8.2E–01 |
|          | CRHR1         | 17: 43699267–43913194            | 86             | 0.09         | 0.18                   | 4.8E–08        | 1.8E–05 | 2.5E–02 | 7.6E–01 |
|          | KANSL1-AS1    | 17: 44270942–44274089            | 27             | 0.27         | 0.43                   | 3.4E–10        | 1.4E–06 | 2.5E–02 | 7.6E–01 |
|          | LRRC37A       | 17: 44370999–44415160            | 75             | 0.27         | 0.20                   | 7.9E–08        | 2.5E–05 | 9.7E–02 | 4.5E–01 |
|          | LRRC37A2      | 17: 44588777–44630815           | 130            | 0.37         | 0.18                   | 3.1E–07        | 8.5E–05 | 8.4E–02 | 5.8E–01 |
| q22      | STXBP4        | 17: 53062975–53241646           | 609            | 0.20         | 0.01                   | 1.4E–25        | 1.6E–25 | 3.3E–02 | 1.5E–06 |
colocalization, not the effect of modeled gene’s expression levels. This additional filter narrowed the number of candidate genes to 901.

### 2.4 Transcriptome-wide association study (TWAS) analyses

Using the functional weights of those 901 genes and summary level GWAS data, we assessed the association between predicted gene expression and breast cancer risk. We performed summary-based imputation using the ImpG-Summary algorithm (Pasaniuc et al., 2014). Briefly, let \( Z \) be a vector of standardized association statistics (\( z \) scores) of SNPs for a trait at a given cis locus, \( \Sigma_{s} \) be the LD matrix from reference genotype data and let \( W = (w_1,w_2,w_3,...,w_j) \) be the weights from the expression prediction model precompiled using the reference panel. Under the null hypothesis that none of the SNPs with \( w_i \neq 0 \) is associated with disease, the test statistic \( w_z/(\Sigma_{s}W)^{1/2} \) follows a normal distribution with mean = 0 and variance = 1. To account for finite sample size and instances where \( \Sigma_{s} \) was not invertible, we adjusted the diagonal of the matrix using a technique similar to ridge regression with \( \lambda = 0.1 \).

### 2.5 Case-only TWAS

To assess whether genetically predicted expression was differentially associated with ER+ and ER- breast cancer, we applied the TWAS procedure described above to the \( Z \) statistics from the BCAC case-only analysis. Following arguments in Barfield et al. (2018) the standard TWAS statistic applied to a case-only GWAS results tests hypothesis \( H_0: \beta_2 - \beta_1 = 0 \). This is similar to a conventional multinomial logistic model for subtype-specific breast cancer risk, with expression log odds ratio \( \beta_2 \) for ER- disease and \( \beta_1 \) for ER+ disease, under which scenario, the expression log odds ratio comparing ER- to ER+ cases is \( \beta_2 - \beta_1 \).

### 2.6 Conditional analyses

Colocalization makes the interpretation of TWAS hits challenging (Mancuso et al., 2019; Wainberg et al., 2019). In addition to the main TWAS analysis, we also performed conditional and joint (COJO) multiple-SNP analysis at each TWAS significant gene location to distinguish colocalization, and to identify gene(s) independently responsible for the statistical association at each locus. COJO approximates the results of a joint conditional analysis including predicted expression levels from multiple proximal genes. The original COJO approach was designed to assess the association of
individual SNPs with a phenotype; we used an extension that jointly models the associations between multiple linear combinations of individual SNPs (Gusev et al., 2016). We conducted two types of COJO: (a) For regions in which multiple associated features were identified (within 500 kb of each other, i.e., colocalization), we jointly modeled these significant TWAS genes to determine the strongest associated gene (or infer independent signals); (b) To provide information on whether the TWAS gene was responsible for the observed SNP-trait association, we also evaluated whether the GWAS-identified index SNPs remained significant after conditioning on the genes within the same region.

3 | RESULTS

3.1 | Breast cancer TWAS

We selected 12,696 transcripts from the 67 GTEx breast tissue samples of European-ancestry women that passed quality control. Based on GCTA-REML analysis, breast-tissue expression levels for 1,355 of these genes were heritable (p value for cis-h2 < .01). We then built linear predictors for these heritable genes and estimated prediction r2 using fivefold cross-validation. A total of 454 genes failed our cross-validation r2 requirement (r2 > .01), and we performed TWAS on the remaining 901 genes. We defined statistical significance for TWAS results as a marginal p < 5.5 × 10−5 (Bonferroni correction controlling the familywise error rate at ≤0.05 for the 901 genes).

First, to compare with previous GWAS findings and to demonstrate the validity of our results, we performed TWAS analysis in overall breast cancer. We identified 30 genes in 18 cytoband regions associated with breast cancer risk (Table 1). Of these regions, 11 (containing 21 genes) were previously reported breast cancer susceptibility loci (harboring one or more GWAS-significant SNP). Five genes in the remaining seven regions were previously reported in TWAS or EUGENE analyses (LINC00886, CTD-2323K18.1, MAN2C1, NUP107, and CPNE1), while the remaining four genes in these regions were novel (MAEA, GDI2,ULK3, and HSD17B1P1). NUP107 and CPNE1 did not pass a stringent Bonferroni significance threshold in Wu et al. (2018) but passed a less-stringent false discovery rate threshold.

We also carried out analyses focusing on breast cancer subtypes. We found 20 genes associated with ER+ breast cancer, and six genes associated with ER− breast cancer (p < .05/901 = 5.5 × 10−5; Table 1). In our results, all genes associated with ER− disease were also associated with ER+ disease, as well as with overall breast cancer risk. Using a more stringent threshold on the strength of the genetic predictor for expression (cross-validation r2 > .1; 383 genes passed this threshold), we found four TWAS significant (p < .05/383 = 1.3 × 10−4) genes for ER− disease, 14 genes for ER+ disease, and 19 genes for overall breast cancer (18 out of 19 genes are included in Table 1 except for one gene, CTD/3110H11.1). As before, these gene sets were nested within each other.

3.2 | Difference of TWAS signal across breast cancer subtypes

We tested whether the imputed gene expression-breast cancer associations differed by subtype using GWAS summary statistics from a case-only analysis, which specifically compared ER+ with ER− breast cancer patients (see Section 2 for details), scanning through...
901 eligible genes. Two genes, HIST2H2BA and STXBP4, showed significant associations (p < .05/901) with ER status among cases (Figure 1). These two genes were associated with ER+ breast cancer but not associated with ER− breast cancer.

### 3.3 GWAS signal conditioning on TWAS gene expression

As shown in Table 2, 21 (of 30) TWAS-significant genes were located near GWAS signals. To examine whether the observed GWAS signal within the gene region could be explained by the expression of that gene, we performed additional analyses conditioning SNP-cancer associations on the predicted expression of that particular significant TWAS gene (See Section 2 and Figure S1, for details). We found that for most regions, GWAS SNPs were no longer associated with the risk of breast cancer once conditioned on the expression of TWAS gene in the region: 15 of 21 genes had no SNPs with a conditional GWAS p value smaller than the genome-wide significant threshold (5 × 10⁻⁸). Thus, there were six genes for which the

| TABLE 2 Summary of conditional analysis at known breast cancer risk region |
|---------------------------------------------------------------|
| **Gene**        | **Before conditional analysis** | **After conditional analysis** | **Smallest conditional p value before and after COJO** |
|-----------------|---------------------------------|-------------------------------|-----------------------------------------------|
|                 | Number of SNPs | Number of significant SNPs | Index GWAS SNP p value | Number of significant SNPs | Index SNP | Smallest conditional p value | Ratio | Magnitude of change in the minimum p value before and after COJO |
| ALS2CR12       | 480 | 12 | 8.2E−17 | 1 | rs3769823 | 6.40E−09 | 0.92 | 1.28E−08 |
| ATG10          | 619 | 24 | 6.9E−13 | 0 | rs891159 | 1.20E−07 | 1.00 | 5.75E−06 |
| ATP6AP1L       | 581 | 24 | 6.9E−13 | 0 | rs891159 | 1.70E−07 | 1.00 | 4.06E−06 |
| CASP8          | 493 | 12 | 8.2E−17 | 6 | rs3769823 | 3.90E−12 | 0.50 | 2.10E−06 |
| CRHR1          | 229 | 13 | 1.5E−10 | 0 | rs17763086 | 1.40E−05 | 1.00 | 1.07E−05 |
| CRHR1-IT1      | 230 | 13 | 1.5E−10 | 0 | rs17763086 | 9.00E−05 | 1.00 | 1.52E−06 |
| HIST2H2BA      | 202 | 19 | 3.5E−52 | 1 | rs11249433 | 7.40E−24 | 0.95 | 4.73E−29 |
| KANSL1-AS1     | 34  | 13 | 1.5E−10 | 0 | rs17763086 | 1.60E−01 | 1.00 | 9.38E−10 |
| L3MBTL3        | 724 | 13 | 1.7E−12 | 0 | rs6569648 | 1.40E−03 | 1.00 | 1.21E−09 |
| LRRC37A        | 285 | 13 | 1.5E−10 | 0 | rs17763086 | 4.60E−05 | 1.00 | 3.26E−06 |
| LRRC37A4P      | 285 | 13 | 1.5E−10 | 0 | rs17763086 | 4.60E−05 | 1.00 | 3.26E−06 |
| MRPL23-AS1     | 557 | 36 | 2.4E−33 | 18 | rs569550 | 1.20E−29 | 0.50 | 2.00E−04 |
| NUDT17         | 112 | 17 | 1.5E−10 | 0 | rs36107432 | 3.90E−05 | 1.00 | 3.85E−06 |
| RP11-15A1.7    | 594 | 32 | 1E−16   | 0 | rs10426528 | 4.60E−07 | 1.00 | 2.17E−10 |
| RP11-250B2.5   | 503 | 8  | 2.7E−09 | 0 | rs9343989 | 1.00E−03 | 1.00 | 2.70E−06 |
| RP11-554A11.9  | 532 | 36 | 2.8E−44 | 33 | rs680618 | 2.80E−44 | 0.08 | 1.00E+00 |
| RP11-73O6.3    | 665 | 13 | 1.7E−12 | 0 | rs6569648 | 1.70E−03 | 1.00 | 1.00E−09 |
| STXB4          | 687 | 46 | 2E−28   | 0 | rs244335 | 2.10E−04 | 1.00 | 9.52E−25 |
| ZNF155         | 597 | 32 | 1E−16   | 5  | rs10426528 | 2.00E−09 | 0.84 | 5.00E−08 |
| ZNF404         | 551 | 32 | 1E−16   | 0  | rs10426528 | 5.30E−07 | 1.00 | 1.89E−10 |
| LRRC37A2       | 152 | 2  | 2E−08   | 0  | rs199498 | 1.80E−02 | 1.00 | 1.11E−06 |

Abbreviations: COJO, conditional and joint; GWAS, genome-wide association studies; SNP, single nucleotide polymorphisms.

aProportion of marginally significant SNPs that are not significant in conditional analyses. Analysis was performed using GWAS summary statistics of ER+ subtypes. The difference between marginal SNP tests for association (GWAS p values) and the SNP p values conditional on significant TWAS genes provides some evidence regarding the independence of the TWAS and single-SNP association signals. The number and proportion of SNPs that are genome-wide significant before and after conditioning on a TWAS-significant gene summarizes the degree single-SNP associations are dependent on (or independent of) the TWAS association.
GWAS SNP remained significantly associated with breast cancer risk at the genome-wide threshold ($5 \times 10^{-8}$) after conditioning on TWAS gene expression. The region containing HIST2H2BA had only one genome-wide significant SNP remaining, and the region containing ZNF155 and ZNF404 had five genome-wide significant SNPs remaining, indicating that the expression of identified genes might explain some but not all of the SNP-breast cancer associations in these regions. For CASP8 and MRPL23-AS1 regions, half of the GWAS hits remained genome-wide significant, and for the RP11-554A11.9 region, 33 out of 36 GWAS SNPs remained (Figures 2 and S1). These results suggest that the genetic association between breast cancer risk and those regions may not be mediated by transcriptional regulation of the genes on which we conditioned.

### 3.4 Mutually adjusting for TWAS-significant genes in the same region

As shown in Table 3, we identified six regions with more than one TWAS-significant gene: 2q33 (CASP8, ALS2CR12), 5q14 (ATG10, ATP6AP1L), 6q22 (RP11-7306.3, L3MBTL3), 15q24 (ULK3, MAN2C1, CTD-2323K18.1), 17q21 (LRRC37A4p, CRHR1-IT1, CRHR1, KANSL1-AS1, LRRC37A, LRRC37A2), and 19q13 (ZNF404, ZNF155, RP11-15A1.7). After mutually conditioning on the predicted expression of all significant genes in the same regions, ten genes remained nominally significant ($p < 0.05$). For some regions, only one gene remained, that is ATG10 for 5q14, L3MBTL3 for 6q22 and CRHR1-IT1 for 17q21 (Figures 3a and S2); while for other regions, multiple genes remained significant, including CASP8 and ALS2CR12 for 2q33.
### TABLE 3  Conditional and joint analysis of gene region with multiple TWAS significant genes

| Region\(^a\) | Gene (colocalized) | Marginal TWAS | COJO | p Value |
|--------------|-------------------|---------------|------|---------|
| 2q33         | ALS2CR12          | 6.7           | 4.6  | 3.70E−06|
|              | CASP8             | −5.22         | −2   | 5.00E−02|
| 5q14         | ATG10             | −6.37         | −6.37| 1.85E−10|
|              | ATP6AP1L          | −5.18         | −0.85| 0.4     |
| 6q22         | RP11-73O6.3       | −6.92         | 0.18 | 0.86    |
|              | L3MBTL3           | −7.46         | 8.45E−14|
| 15q24        | ULK3              | −4.11         | −4.1 | 3.90E−05|
|              | MAN2C1            | −4.95         | −5   | 7.40E−07|
|              | CTD-2323K18.1     | −4.95         | 7.49E−07|
| 17q21        | LRRC37A4P         | 6.29          | 0.25 | 0.8     |
|              | CRHR1-IT1         | −6.3          | −6.3 | 2.91E−10|
|              | CRHR1             | −5.46         | −0.28| 0.78    |
|              | KANSL1-AS1        | −6.28         | −0.04| 0.97    |
|              | LRRC37A           | −5.37         | 1.83 | 0.07    |
|              | LRRC37A2          | −5.12         | 1.81 | 0.07    |
| 19q13        | ZNF404            | 7.35          | 3.5  | 0.001   |
|              | ZNF155            | 5.75          | 2.8  | 0.042   |
|              | RP11-15A1.7       | 6.81          | 2.8  | 0.005   |

Abbreviations: COJO, conditional and joint; TWAS, transcriptome-wide association studies.

\(^a\)Bolded genes remain significant in conditional analyses. Analysis was performed using GWAS summary statistics of ER+ subtypes. Our primary goal in these analyses is to establish whether any of the marginally significant TWAS genes remains significant after conditioning for the most significant gene in the region; since all of the regions with multiple significant genes contain 2–3 significant genes, using a conditional p value threshold of .05 is a reasonable threshold for identifying independent signals.

ULK3 and MAN2C1 for 15q24, and ZNF404, ZNF155, and RP11-15A1.7 for 19q13 (Figures 3b and S2).

### 4  DISCUSSION

We conducted a TWAS analysis using GTEx mammary tissue gene expression data and GWAS summary data from the largest meta-analysis for breast cancer risk. We assessed associations between overall breast cancer risk and ER+ versus ER− disease. We found 30 genes significantly associated with overall breast cancer risk, 20 genes associated with the ER+ subtype, and six genes with the ER− subtype.

These results are consistent with previous reports from TWAS or similar gene-based approaches, which used various algorithms to build gene expression models. For example, of the 30 genes that we found significantly related to overall breast cancer risk, 23 were also significant in Wu et al. (2018) with very similar test statistics (correlation = 0.96 for the z scores between our and Wu’s results), and six were significant in Ferreira et al. (2019).

One of the six genes we classified as significantly associated with ER− breast cancer was also found significantly associated with ER− breast cancer in Ferreira et al. (2019). Among these studies, the approach taken by Wu et al. was the most similar to ours. Only seven of the 30 genes that we identified were not identified by Wu et al. (2018), probably due to different cis-SNP selection criteria and different candidate genes selected for testing. We defined cis-SNPs using a 500 KB window around the gene boundary and included only candidate genes with a significant heritability, while Wu et al. used a 2 MB cis-SNP window and included genes with a prediction performance of at least 0.01 without heritability filtering. For genes whose expression could not be predicted well, Wu et al. built models using only SNPs located in promoter or enhancer regions. Despite these methodological differences, the two TWAS results were highly concordant. However, we did not replicate any of the findings in Hoffman et al. (2017) and Gao et al. (2017), which may reflect the smaller sample size of the breast cancer GWAS used in their analyses (3,370 cases and 19,717 controls in Hoffman et al.; 10,597 overall breast cancer cases, 3,879 ER− cases and 11,358 controls in Gao et al.). Specifically, three of the previously reported genes were excluded by our stringent
QC procedure (DHODH, ANKLE1 from Hoffman et al. and TP53INP2 from Gao et al. were not heritable in our analysis) and one was not significant in our analysis (RCCD1 from Hoffman et al. \( p = .0032 \) for overall breast cancer). Both Hoffman et al. and Gao et al. used GWAS results based on a mixed population of European, African, and Asian ancestry (which shared a small set of European samples with our GWAS: \( N < 5,700 \) individuals from CGEMS and the BPC3, less than 2% of our GWAS sample). They also used different tissues to build their prediction weights: overall breast tissue (men and women combined, all ethnicities) and whole blood tissue (men and women combined, European ancestry).

Of the 30 genes associated with breast cancer risk in our study, 21 fell into known GWAS regions whereas nine were not close to any known GWAS hit and were, therefore, considered novel. Of these nine genes, five were identified and discussed in Wu et al. (2018) or Ferreira et al. (2019). The four genes uniquely identified in the present study were GDI2, HSD17B1P1, MAEA, and ULK3, several of which have been reported to play a role in breast tumorigenesis or related biological processes. For example, the expression of GDI2 has been linked with breast cancer through its contribution to enhanced epidermal growth factor receptor endocytosis (EGFR; de Graauw et al., 2014). HSD17B1P1 is a pseudo-gene related to HSD17, which participates in steroid hormone biosynthesis, metabolism, and signaling pathways potentially related to breast cancer risk (Jakubowska et al., 2010). These findings lend support to our results and suggested that further investigation into the roles of the novel genes identified for breast cancer is required.

We performed several conditional analyses not reported in previous TWAS. We examined the local GWAS signals conditioning on the expression of TWAS genes, to provide a measure of how well the expression level of identified TWAS genes explained the local GWAS signals. For many loci, these genes explained a large proportion

---

**Figure 3** COJO for regions with multiple TWAS associations. For each plot, the top panel shows all genes in the locus. After COJO analysis, the marginally associated genes are highlighted in blue, while those that remain jointly significant are highlighted in green (in this case, L3MBTL3, CASP8, and ALS2CR12). The bottom panel shows a Manhattan plot of the GWAS signals before (gray) and after (blue) conditioning on the significant (green) genes. (a) COJO results for 6q22 (only one gene remains significant after COJO). (b) COJO results for 2q33 (an example of multiple genes remaining jointly significant after COJO). COJO, conditional and joint analysis; GWAS, genome-wide association studies; TWAS, transcriptome-wide association studies.
of the local GWAS signals and were thus candidates for downstream experimental validation. We also identified candidate genes driving the statistical associations in regions with more than one TWAS gene (usually also regions with known GWAS risk loci) by jointly modeling multiple nominally significant genes. For example, previous studies have suggested that polymorphisms in CASP8 are associated with breast cancer risk (Cox et al., 2007), whereas a recent paper has shown that the most significant signal in this region is for the imputed intronic SNP rs1830298 in ALS2CR12 (telomeric to CASP8; Lin et al., 2015). Our results provide clarification on whether CASP8 or ALS2CR12 expression were more strongly associated with breast cancer risk, since both genes remained significantly associated with breast cancer risk after conditioning on the expression of the other (the conditional p value for ALS2CR12 was $3.70 \times 10^{-6}$), whereas the conditional p value for CASP8 was .05).

Eleven of the 12 GWAS hits disappeared after adjusting for the expression of ALS2CR12, while half of the GWAS hits remained after adjusting for the expression of CASP8. Therefore, we believe that ALS2CR12 SNPs have a stronger effect and are associated with breast cancer through ALS2CR12 expression, while CASP8 remains an additional independent hit, consistent with the latest fine-mapping results (Lin et al., 2015).

Because the genes found to be associated with ER− disease were also associated with ER+ disease, and these, in turn, were associated with overall breast cancer risk, it is difficult to conclude whether the differences in gene sets are due to distinct mechanisms underlying breast cancer subtypes or due to a lack of statistical power because of the smaller disease subtype sample sizes. To address this question, we further incorporated a case-only TWAS comparing ER+ versus ER− breast cancer. We identified two genes, STXBP4 and HIST2H2BA, associated with ER status, which were significantly associated only with ER+ but not ER− breast cancer. Previous studies supported the link between rs6504950 (a SNP in STXBP4) and overall breast cancer risk (Antoniou et al., 2010; Warren Andersen et al., 2013). It has also been hypothesized that the risk allele for the two top breast cancer candidate SNPs, rs2787486 and rs244353, affected gene expression of STXBP4 (Darabi et al., 2016) and CD4 memory cells (Hnisz et al., 2013). One potential explanation for the association between STXBP4 and breast cancer risk is that it encodes syntaxin binding protein 4, a scaffold protein. In addition, STXBP4 functions to stabilize and degrade TP63 isoform (a member of the TP53 tumor suppressor protein family), a biologically plausible candidate cancer susceptibility gene. Similarly, SNPs rs2580520 and rs11249433 upstream of HIST2H2BA have been identified as breast cancer susceptibility alleles in a previous GWAS (Bogdanova, Helbig, & Dörk, 2013). Our results suggest that functional and pathway analyses targeting these two genes are likely to shed new light on the differences in tumorigenesis and progression mechanisms between ER+ and ER− patients.

By building gene expression linear predictors in GTEx breast tissue, our analysis offers a tissue-specific model of gene expression. The gene regulatory mechanisms in female breast tissue are arguably the most suitable for studying breast cancer. Moreover, by restricting our reference population to women of European ancestry, rather than mixing genders and ancestries, the resulting gene expression model was a better match to our breast cancer GWAS summary statistics. By using the largest GWAS meta-analysis currently available, we greatly improved the power compared with previous work by Hoffman et al. (2017) and Gao et al. (2017). Finally, by using case-only GWAS summary statistics, we provided insights into genes associated with breast cancer subtype specific risk compared with Wu et al. (2018) and Ferreira et al. (2019).

Similar to previous work by Wu et al. (2018) and Hoffman et al. (2017), our analyses focused on genetic tools trained using expression from breast tissue, chosen because of its direct relevance to breast carcinogenesis. However, given the relatively small sample size in the breast tissue eQTL panel, this choice limited both our power to detect genes with cis-heritable expression and the precision of estimated genetic predictors for heritable transcripts. The genetic regulation of expression is constant across tissues for many genes, suggesting that considering other tissues with larger eQTL sample sizes or combining eQTL evidence across tissues may improve power. In addition, other tissues may be relevant for breast cancer development. For example, considering that obesity and hormonal signaling have been linked to breast cancer risk (Bertolini, 2013), gene expression in adipose tissue and brain tissue may have parallel involvement with breast cancer etiology. We are currently developing methods for cross-tissue TWAS, using sCCA (sparse canonical correlation analysis) to build features that combine gene expression values across tissues that share similar genetic regulation mechanisms, while allowing tissues with different regulation patterns to contribute to different features (Feng, Pasaniuc, Major, & Kraft, 2018).

In conclusion, we have identified new breast cancer target genes both for functional experiments and as causal gene candidates in the significant TWAS gene regions. We have also identified associations between gene expression and breast cancer risk specific to disease subtypes, where two novel genes have been found specifically associated with ER+ breast cancer risk.
This analytic strategy warrants application in studies aimed at defining the genomic architecture of cancers other than breast cancer.

ACKNOWLEDGMENTS
The authors thank the Cellex Foundation for providing research facilities and equipment.

The breast cancer genome-wide association (BCAC) is funded by Cancer Research UK (C1287/A16563, C1287/A10118), the European Union’s Horizon 2020 Research and Innovation Programme (grant nos. 634935 and 633784 for BRIDGES and B-CAST, respectively), and by the European Community’s Seventh Framework Programme under grant agreement number 223175 (grant no. HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Genotyping of the OncoArray was funded by the NIH grant U19 CA148065, and Cancer UK grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344), and the Ministère de l’Économie, Science et Innovation du Québec through Genome Québec and the PSRIIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community’s Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175; COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, and C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (U19 CA148537, 1U19 CA148065, and 1U19 CA148122—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065.

The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute. The Australian Breast Cancer Family Study was partly funded by ELAN-Fond of the University Hospital of Erlangen. The ABCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BCFR was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). For the BCFR-ny, BCFR-pa, BCFR-ut this work was supported by grant UMI CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. For BIGGS, E. S. is supported by NIHR Comprehensive Biomedical Research Centre, Guy’s and St. Thomas’ NHS Foundation Trust in partnership with King’s College London, United Kingdom. I. T. is supported by the Oxford Biomedical Research Centre. B. O. C. S. is supported by funds from Cancer Research UK (C8620/A8372/A15106) and the Institute of Cancer Research (UK). B. O. C. S. acknowledges NHS funding to the Royal Marsden/Institute of Cancer Research NIHR Specialist Cancer Biomedical Research Centre. The BREast Oncology Galician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Conselleria de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de...
Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain (grant EC11-192). Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economía y Competitividad, Xunta de Galicia, Spain. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). The CAMA study was funded by Consejo Nacional de Ciencia y Tecnología (CONACyT; SALUD-2002-C01-7462). Sample collection and processing were funded in part by grants from the National Cancer Institute (NCI R01CA120120 and K24CA169004). C. B. C. S. is funded by the Canadian Cancer Society (grant no. 313404) and the Canadian Institutes of Health Research. C. C. G. P. is supported by funding from the University of Crete. The CECILE study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale Contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l’Alimentation, de l’Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The CNIO-BCS was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). COLBCCC is supported by the German Cancer Research Center (DKFZ), Heidelberg, Germany. Diana Torres was in part supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation. The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The CTS was initially supported by the California Breast Cancer Act of 1993 and the California Breast Cancer Research Fund (contract 97-10500) and is currently funded through the National Institutes of Health (R01 CA77398, UM1 CA164917, and U01 CA199277). The collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. HAC receives support from the Lon V Smith Foundation (LVS39420). The University of Westminster curates the DietCompLyf database funded by Against Breast Cancer Registered Charity No. 1121258 and the NCRN. The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l’Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). FHRISK is funded from NIHR grant PgfAR 0707-10031. The GC-HBOC (German Consortium of Hereditary Breast and Ovarian Cancer) is supported by the German Cancer Aid (grant no 110837, coordinator: Rita K. Schmutzler, Cologne). This work was also funded by the European Regional Development Fund and Free State of Saxony, Germany (LIFE—Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. The GEPARSIXTO study was conducted by the German Breast Group GmbH. The GESBC was supported by the German Cancer Aid (DKFZ). GLACIER was supported by Breast Cancer Now, CRUK and Biomedical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. The HABCS study was supported by the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. The HEBCS was financially supported by the Helsinki University Hospital Research Fund, the Finnish Cancer Society, and the Sigrid
Research Fellowship. The KOHBRA study was partially supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (HI16C1127; 1020350; 1420190). L. A. A. B. C. is supported by grants (1RB-0287, 3PB-0102, 5PB-0018, and 10PB-0098) from the California Breast Cancer Research Program. Incident breast cancer cases were collected by the USC Cancer Surveillance Program (CSP) which is supported under subcontract by the California Department of Health. The CSP is also part of the National Cancer Institute’s Division of Cancer Prevention and Control Surveillance, Epidemiology, and End Results Program, under contract number N01CN25403. L. M. B. C. is supported by the “Stichting tegen Kanker.” D. L. is supported by the FWO. The MABCs study is funded by the Research Centre for Genetic Engineering and Biotechnology “Georgi D. Efremov” and supported by the German Academic Exchange Program, DAAD. The MARIE study was supported by the Deutsche Krebshilfe e.V. (70-2892-BR 1, 106332, 108253, 108419, 110826, 110828), the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany (01KH0402). MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1,000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects “5 × 1,000”). The MCBCS was supported by the NIH grants CA192393, CA116167, CA176785, and NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414, and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.” The MEC was supported by NIH grants CA63464, CA54281, CA098758, CA132839, and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931,
CA116201, CA140286, and CA177150. M. S. K. C. C. is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. The work of MTLGEBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the “CIHR Team in Familial Risks of Breast Cancer” program—grant no. CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade—grant no. PSR-SIIIRI-701. MYBRCA is funded by research grants from the Malaysian Ministry of Higher Education (UM.C/HIR/ MOHE/06) and Cancer Research Malaysia. MYMAMMO is supported by research grants from Yayasan Sime Darby LPGA Tournament and Malaysian Ministry of Higher Education (RP046B-15HTM). The NBCS has been supported by the Research Council of Norway grant 193387/V50 (to A.-L. B.-D. and V. N. K.) and grant 193387/H10 (to A.-L. B.-D. and V. N. K.). South-Eastern Norway Health Authority (grant 39346 to A.-L.B-D. and 27208 to V. N. K.) and the Norwegian Cancer Society (to A.-L. B.-D. and 419616-71248-PR-2006-0282 to V. N. K.).

It has received funding from the K.G. Jebsen Centre for Breast Cancer Research (2012-2015). The NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The Carolina Breast Cancer Study was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, and U01 CA179715), and the North Carolina University Cancer Research Fund. The NGOBCS was supported by Grants-in-Aid for the Third Term Comprehensive Ten-Year Strategy for Cancer Control from the Ministry of Health, Labor and Welfare of Japan, and for Scientific Research on Priority Areas, 17015049 and for Scientific Research on Innovative Areas, 221S0001, from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. The OBCS was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant no. 250083, 122715 and Center of Excellence grant no. 251314), the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation and the special Governmental EVO funds for Oulu University Hospital-based research activities. The ORIGO study was supported by the Dutch Cancer Society (RL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. The POSH study is funded by Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, C1275/A15956, and Breast Cancer Campaign 2010PR62, 2013PR04). PROCAS is funded from NIH grant PGiAR 0707-1003). The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The SABAC study was supported by funding from the Agency for Science, Technology, and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The SBCGS was supported primarily by NIH grants R01CA64277, R01CA148667, UMCA182910, and R37CA70867. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The scientific development and funding of this project were, in part, supported by the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network U19 CA148065. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank. The SCCS is supported by a grant from the National Institutes of Health (R01 CA092447). Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from the National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC).
Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry. SEARCH is funded by Cancer Research UK (C490/A10124 and C490/A16561) and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. SEBCS was supported by the BRL (Basic Research Laboratory) program through the National Research Foundation of Korea funded the BRL (Basic Research Laboratory) program through Clinical Academic Reserve. SEBCS was supported by the University of Cambridge has received salary support for the UK National Institute for Health Research Biomedi-Research Centre at the University of Cambridge. The UKNCR Study was supported by Imbral Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The US3SS study was supported by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The WAABCS study was supported by grants from the National Cancer Institute of the National Institutes of Health (R01 CA89085 and P50 CA125183 and the D43 TW009112 grant), Susan G. Komen (SAC110026), Dr. Ralph and Marian Falk Medical Research Trust, and the Avon Foundation for Women. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100004C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01. D. G. E. is supported by the all Manchester NIHR Biomedical research center Manchester (IS-BRC-1215-2007), HUNBOCS, Hungarian Breast, and Ovarian Cancer Study were supported by Hungarian Research Grant KTIA-OTKA CK-80745, NKFI_OTKA K-112228. C. I. received support from the Survey, Recruitment, and Biospecimen Resource at Georgetown University (NIH/NCI P30-CA-51008) and the Jess and Mildred Fisher Center for Hereditary Cancer and Clinical Genomics Research. K. M. is supported by CRUK C18281/A19169. City of Hope Clinical Cancer Community Research Network and the Hereditary Cancer Research Registry, supported in part by Award Number RC4CA153828 (PI: J Weitzel) from the National Cancer Institute and the office of the Director, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The ICO study is supported by the Asociación Española Contra el Cáncer (AECC), The Instituto de Salud Carlos III (organismo adscrito al Ministerio de Economía y Competitividad) and “Fondo Europeo de Desarrollo Regional (FEDER), una manera de hacer Europa” (P10/01422, P113/00285, PIE13/00022, PI15/00854, PI16/00563, and CIBERONC) and The Institut Català de la Salut and Autonomous Government of Catalonia (2009SGR290, 2014SGR338 and PERIS Project MedPerCan). Dr. Beth Karlan is funded by the American Cancer Society Early Detection Professorship and health information sources who have contributed to the study. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The UKOPS study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The US3SS study was supported by Massachusetts (K. M. E., R01CA47305), Wisconsin (P. A. N., R01 CA47147), and New Hampshire (L. T.-E., R01CA69664) centers, and Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The US3SS study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The WAABCS study was supported by grants from the National Cancer Institute of the National Institutes of Health (R01 CA89085 and P50 CA125183 and the D43 TW009112 grant), Susan G. Komen (SAC110026), Dr. Ralph and Marian Falk Medical Research Trust, and the Avon Foundation for Women. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100004C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01. D. G. E. is supported by the all Manchester NIHR Biomedical research center Manchester (IS-BRC-1215-2007), HUNBOCS, Hungarian Breast, and Ovarian Cancer Study were supported by Hungarian Research Grant KTIA-OTKA CK-80745, NKFI_OTKA K-112228. C. I. received support from the Survey, Recruitment, and Biospecimen Resource at Georgetown University (NIH/NCI P30-CA-51008) and the Jess and Mildred Fisher Center for Hereditary Cancer and Clinical Genomics Research. K. M. is supported by CRUK C18281/A19169. City of Hope Clinical Cancer Community Research Network and the Hereditary Cancer Research Registry, supported in part by Award Number RC4CA153828 (PI: J Weitzel) from the National Cancer Institute and the office of the Director, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The ICO study is supported by the Asociación Española Contra el Cáncer (AECC), The Instituto de Salud Carlos III (organismo adscrito al Ministerio de Economía y Competitividad) and “Fondo Europeo de Desarrollo Regional (FEDER), una manera de hacer Europa” (P10/01422, P113/00285, PIE13/00022, PI15/00854, PI16/00563, and CIBERONC) and The Institut Català de la Salut and Autonomous Government of Catalonia (2009SGR290, 2014SGR338 and PERIS Project MedPerCan). Dr. Beth Karlan is funded by the American Cancer Society Early Detection Professorship and health information sources who have contributed to the study. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The UKOPS study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The US3SS study was supported by Massachusetts (K. M. E., R01CA47305), Wisconsin (P. A. N., R01 CA47147), and New Hampshire (L. T.-E., R01CA69664) centers, and Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The US3SS study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The WAABCS study was supported by grants from the National Cancer Institute of the National Institutes of Health (R01 CA89085 and P50 CA125183 and the D43 TW009112 grant), Susan G. Komen (SAC110026), Dr. Ralph and Marian Falk Medical Research Trust, and the Avon Foundation for Women. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100004C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01. D. G. E. is supported by the all Manchester NIHR Biomedical research center Manchester (IS-BRC-1215-2007), HUNBOCS, Hungarian Breast, and Ovarian Cancer Study were supported by Hungarian Research Grant KTIA-OTKA CK-80745, NKFI_OTKA K-112228. C. I. received support from the Survey, Recruitment, and Biospecimen Resource at Georgetown University (NIH/NCI P30-CA-51008) and the Jess and Mildred Fisher Center for Hereditary Cancer and Clinical Genomics Research. K. M. is supported by CRUK C18281/A19169. City of Hope Clinical Cancer Community Research Network and the Hereditary Cancer Research Registry, supported in part by Award Number RC4CA153828 (PI: J Weitzel) from the National Cancer Institute and the office of the Director, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The ICO study is supported by the Asociación Española Contra el Cáncer (AECC), The Instituto de Salud Carlos III (organismo adscrito al Ministerio de Economía y Competitividad) and “Fondo Europeo de Desarrollo Regional (FEDER), una manera de hacer Europa” (P10/01422, P113/00285, PIE13/00022, PI15/00854, PI16/00563, and CIBERONC) and The Institut Català de la Salut and Autonomous Government of Catalonia (2009SGR290, 2014SGR338 and PERIS Project MedPerCan). Dr. Beth Karlan is funded by the American Cancer Society Early Detection Professorship
(SIOP-06-258-01-COUN) and the National Center for Advancing Translational Sciences (NCATS), grant UL1TR000124. A.V. is supported by the Spanish Health Research Foundation, Instituto de Salud Carlos III (ISCIII), partially supported by FEDER funds through Research Activity Intensification Program (contract grant nos. INT15/00070, INT16/00154, INT17/00133), and through Centro de Investigación Biomédica en Red de Enfermedades Raras CIBERER (ACCI 2016: ER17PIAC7112/2018); Autonomous Government of Galicia (Consolidation and structuring program: IN607B), and by the Fundación Mutua Madrileña (call 2018). The GEMO resource was initially funded by the French National Institute of Cancer (INCa, PHRC Ile de France, grant AOR 01 082, 2001–2003, grant 2013-1-BCB-01-ICH-1), the Association “Le cancer du sein, parlons-en!” Award (2004) the Association for International Cancer Research (2008–2010), and the Fondation ARC pour la recherche sur le cancer (grant PJA 20151203365). It also received support from the Canadian Institute of Health Research for the “CIHR Team in Familial Risks of Breast Cancer” program (2008–2013), and the European commission FP7, Project «Collaborative Ovarian, breast and prostate Gene-environment Study (COGS), Large-scale integrating project» (2009–2013). G. E. M. O. is currently supported by the INCa grant SHS-E-SP 18-015.

OSUCC was funded by the Ohio State University Comprehensive Cancer Center. Leigha Senter, Kevin Sweet, Caroline Craven, Julia Cooper, Amber Aiels, and Michelle O’Conor aided in the recruitment of BRCA1/2 study participants and data collection. Robert Pilarski aided in recruitment and data collection of TNBC cases from the Stefanie Spielman Breast Bank.

Clinical Genetics Branch, NCI: the Intramural Research Program of the US National Cancer Institute, NIH, Division of Cancer Epidemiology and Genetics, and by support services contracts NO2-CP-11019-50, NO2-CP-21013-63 and NO2-CP-65504 with Westat, Inc, Rockville, MD.

ILUH was funded by the Icelandic Association “Walking for Breast Cancer Research” and by the Landspitali University Hospital Research Fund.

The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON) consists of the following Collaborating Centers: Netherlands Cancer Institute (coordinating center), Amsterdam, NL: M.A. Rookus, F.B.L. Hogervorst, F.E. van Leeuwen, M.A. Adank, M.K. Schmidt, D.J. Jenner; Erasmus Medical Center, Rotterdam, NL: J.M. Colée, A.M.W. van den Ouweland, M.J. Hooning, I.A. Boere; Leiden University Medical Center, NL: C.J. van Asperen, P. Devilee, R.B. van der Luijt, T.C.T.E.F. van Cronenburg; Radboud University Nijmegen Medical Center, NL: M.R. Wevers, A.R. Mensenkamp; University Medical Center Utrecht, NL: M.G.E.M. Ausems, M.J. Koudijs; Amsterdam Medical Center, NL: E.J. Meijers-Heijboer, T.A.M. van Os; VU University Medical Center, Amsterdam, NL: K. van Engelen, J.P.P. Gille; Maastricht University Medical Center, NL: E.B. Gómez-Garcia, M.J. Blok, M. de Boer; University of Groningen, NL: J.C. Oosterwijk, A.H. van der Hout, M.J.E. Mourits, G.H. de Bock; The Netherlands Comprehensive Cancer Organisation (IKNL): S. Siesling, J. Verloop; The nationwide network and registry of histo- and cyto-pathology in the Netherlands (PALGA): E.C. van den Broek. HEBON thanks the study participants and the registration teams of IKNL and PALGA for part of the data collection.

The HEBON study is supported by the Dutch Cancer Society grants NKI1998-1854, NKI2004-3088, NKI2007-3756, the Netherlands Organisation of Scientific Research grant NWO 91109024, the Pink Ribbon grants 110005 and 2014-187.WO76, the BBMRI grant NWO 184.021.007/CP46, and the Transcan grant JTC 2012 Cancer 12-054.

N.N. Petrov Institute of Oncology is supported by the Russian Foundation for Basic Research (grants 17-00-00171, 18-515-45012 and 19-515-25001).

CONFLICT OF INTERESTS
M. W. B. conducts research funded by Amgen, Novartis, and Pfizer. P. A. F. conducts research funded by Amgen, Novartis, and Pfizer. He received Honoraria from Roche, Novartis and Pfizer.

DATA AVAILABILITY STATEMENT
The GWAS summary data that support the findings of overall, ER−, ER+ breast cancer results are openly available in the Breast Cancer Association Consortium (BCAC). The GWAS summary data that support the findings of case-only breast cancer results will also be made available at the BCAC website. http://bcac.cgge.medschl.cam.ac.uk/bcadata/oncoarray/gwas-icogs-and-oncoarray-summary-results/ (Michailidou et al., 2017, pp. 92–94).

REFERENCES
Amos, C. I., Dennis, J., Wang, Z., Byun, J., Schumacher, F. R., Gayther, S. A., ... Easton, D. F. (2016). The OncoArray Consortium: A network for understanding the genetic architecture of common cancers. Cancer Epidemiology, Biomarkers & Prevention, https://doi.org/10.1158/1055-9965.EPI-16-0106
Antoniou, A. C., Beesley, J., McGuffog, L., Sinilnikova, O. M., Healey, S., Neuhausen, S. L., ... Cimba (2010). Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: Implications for risk prediction. Cancer Research, 70(23), 9742–9754. https://doi.org/10.1158/0008-5472.CAN-10-1907
Antoniou, A. C., Goldgar, D. E., Andrieu, N., Chang-Claude, J., Brohet, R., Rookus, M. A., & Easton, D. F. (2005). A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. *Genetic Epidemiology, 29*(1), 1–11. https://doi.org/10.1002/gepi.20074

Aitchley, D. P., Albarracin, C. T., Lopez, A., Valero, V., Amos, C. I., Gonzalez-Angulo, A. M., ... Arun, B. K. (2008). Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *Journal of Clinical Oncology, 26*(26), 4282–4288. https://doi.org/10.1200/JCO.2008.16.6231

Barfield, R., Feng, H., Gusev, A., Wu, L., Zheng, W., Pasaniuc, B., & Kraft, P. (2018). Transcriptome-wide association studies accounting for colocalization using Egger regression. *Genetic Epidemiology, 42*(5), 418–433. https://doi.org/10.1002/gepi.22131

Barnes, D. R., Lee, A., EMBRACE Investigators, kConFab Investigators, Easton, D. F., & Antoniou, A. C. (2012). Evaluation of association methods for analysing modifiers of disease risk in carriers of high-risk mutations. *Genetic Epidemiology, 36*(3), 274–291. https://doi.org/10.1002/gepi.21620

Beggs, A. D., & Hodgson, S. V. (2009). Genomics and breast cancer: The different levels of inherited susceptibility. *European Journal of Human Genetics, 17*(7), 855–856. https://doi.org/10.1038/ejhg.2008.235

Bertolini, F. (2013). Adipose tissue and breast cancer progression: A link between metabolism and cancer. *Breast, 22*(Suppl 2), S48–S49. https://doi.org/10.1016/j.breast.2013.07.009

Blows, F. M., Driver, K. E., Schmidt, M. K., Broeks, A., van Leeuwen, F. E., Wesseling, J., ... Huntsman, D. (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 Medicine, 7*(5):e1000279. https://doi.org/10.1371/journal.pmed.1000279

Bogdanova, N., Helbig, S., & Dörk, T. (2013). Hereditary breast cancer: Ever more pieces to the polygenic puzzle. *Hereditary Cancer in Clinical Practice, 11*(1), 12. https://doi.org/10.1186/1897-4287-11-12

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 68*(6), 394–424. https://doi.org/10.3322/caac.21492

Cai, Q., Zhang, B., Sung, H., Low, S.-K., Kweon, S.-S., Lu, W., ... Zheng, W. (2014). Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nature Genetics, 46*(8), 886–890. https://doi.org/10.1038/ng.3041

Cox, A., Dunning, A. M., Garcia-Closas, M., Balasubramanian, S., Reed, M. W. R., Pooley, K. A., ... Johnson, N., Breast Cancer Association Consortium. (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics, 39*(3), 352–358. https://doi.org/10.1038/ng1981

Darabi, H., Beesley, J., Droit, A., Kar, S., Nord, S., Moradi Marjanneh, M., ... Dunning, A. M. (2016). Fine scale mapping of the 17q22 breast cancer locus using dense SNPs, genotyped within the Collaborative Oncological Gene-Environment Study (COGs). *Scientific Reports, 6*, 32512. https://doi.org/10.1038/srep32512

Delaneau, O., Marchini, J., & Zagury, J.-F. (2011). A linear complexity phasing method for thousands of genomes. *Nature Methods, 9*(2), 179–181. https://doi.org/10.1038/nmeth.1785

Dunning, A. M., Michailloud, K., Kuchenbaecker, K. B., Thompson, D., French, J. D., Beesley, J., ... Edwards, S. L. (2016). Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nature Genetics, 48*(4), 374–386. https://doi.org/10.1038/ng.3521

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D. P., Thompson, D., Ballinger, D. G., ... Ponder, B. A. J. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature, 447*(7148), 1087–1093. https://doi.org/10.1038/nature05887

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57–74. https://doi.org/10.1038/nature11247

Feng, H., Pasaniuc, B., Major, M., & Kraft, P. (2018). Sparse canonical correlation analysis (sCCA) significantly improves power of cross-tissue transcriptome-wide association studies (TWAS). *Genetic Epidemiology, 42*, 698.

Ferreira, M. A., Gamazon, E. R., Al-Ejeh, F., Aittomäki, K., Andrulis, I. L., Anton-Culver, H., ... Chenewx-Trench, G. (2019). Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. *Nature Communications, 10*(1), 1741. https://doi.org/10.1038/s41467-018-08053-5

Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., ... Im, H. K. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics, 47*(9), 1091–1098. https://doi.org/10.1038/ng.3367

Gao, G., Pierce, B. L., Olopade, O. I., Im, H. K., & Huo, D. (2017). Trans-ethnic predicted expression genome-wide association analysis identifies a gene for estrogen receptor-negative breast cancer. *PLOS Genetics, 13*(9):e1006727. https://doi.org/10.1371/journal.pgen.1006727

Garcia-Closas, M., Couch, F. J., Lindstrom, S., Michailloud, K., Schmidt, M. K., Brook, M. N., ... Kraft, P. (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nature Genetics, 45*(4), 392–398, 398e1-2. https://doi.org/10.1038/ng.2561

de Graauw, M., Cao, L., Winkel, L., van Miltenburg, M. H. a M., le Dévédec, S. E., Klop, M., ... van de Water, B. (2014). Annexin A2 deletion delays EGFR endocytic trafficking via coflin activation and enhances EGFR signaling and metastasis formation. *Oncogene, 33*(20), 2610–2619. https://doi.org/10.1038/onc.2013.219

GTEx Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation across human tissues. *Science, 348*(6235), 648–660. https://doi.org/10.1126/science.1262110

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., ... Pasaniuc, B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics, 48*(3), 245–252. https://doi.org/10.1038/ng.3506
Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Mavaddat, N., Antoniou, A. C., Easton, D. F., & Garcia Mancuso, N., Shi, H., Goddard, P., Kichaev, G., Gusev, A., & Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nature Genetics, 51(4), 675–682. https://doi.org/10.1038/s41588-019-0367-1

Mavaddat, N., Shi, H., Goddard, P., Kichaev, G., Gusev, A., & Pasaniuc, B. (2017). Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. American Journal of Human Genetics, 100(3), 473–487. https://doi.org/10.1016/j.ajhg.2017.01.031

Mavaddat, N., Antoniou, A. C., Easton, D. F., & Garcia-Closas, M. (2010). Genetic susceptibility to breast cancer. Molecular Oncology, 4(3), 174–191. https://doi.org/10.1016/j.molonc.2010.04.011

Mavaddat, N., Pharoah, P. D. P., Michailidou, K., Tyer, J., Brook, M. N., Bolla, M. K., ... Garcia-Closas, M. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. Journal of the National Cancer Institute, 107(5), https://doi.org/10.1093/jnci/djv036

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., ... Mahajan, A. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics, 48(10), 1279–1283. https://doi.org/10.1038/ng.3643

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., ... Easton, D. F. (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nature Genetics, 47(4), 373–380. https://doi.org/10.1038/ng.3242

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., ... Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. Nature, 551(7678), 92–94. https://doi.org/10.1038/nature24284

Milne, R. L., Kuchenbaecker, K. B., Michailidou, K., Beesley, J., Kar, S., Lindström, S., ... Simard, J. (2017). Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nature Genetics, 49(12), 1767–1778. https://doi.org/10.1038/ng.3785

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., & Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. PLOS Genetics, 6(4):e1000888. https://doi.org/10.1371/journal.pgen.1000888

Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., ... Price, A. L. (2014). Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics, 30(20), 2906–2914. https://doi.org/10.1093/bioinformatics/btu416

Seal, S., Thompson, D., Renwick, A., Elliott, A., Kelly, P., Barfoot, R., ... Rahman, N. (2006). Truncating mutations in the Fanconi anemia J gene BRF1 are low-penetrance breast cancer susceptibility alleles. Nature Genetics, 38(11), 1239–1241. https://doi.org/10.1038/ng1902

Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., ... Easton, D. F. (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. Nature Genetics, 42(6), 504–507. https://doi.org/10.1038/ng.586

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A. N., Knowles, D. A., Golan, D., ... Kundaje, A. (2019). Opportunities and challenges for transcriptome-wide association studies. Nature Genetics, 51(4), 592–599. https://doi.org/10.1038/s41588-019-0385-z

Warren Andersen, S., Trentham-Dietz, A., Gangnon, R. E., Hampton, J. M., Figueroa, J. D., Skinner, H. G., ... Newcomb, P. A. (2013). The associations between a polygenic score, reproductive and menstrual risk factors and breast cancer risk. Breast Cancer Research and Treatment, 140(2), 427–434. https://doi.org/10.1007/s10549-013-2646-3

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Wu, L., Shi, W., Long, J., Guo, X., Michailidou, K., Beesley, J., ... Zheng, W. (2018). A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nature Genetics, 50(7), 968–978. https://doi.org/10.1038/s41588-018-0132-x

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., ... Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. Nature Genetics, 42(7), 565–569. https://doi.org/10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. American Journal of Human Genetics, 88(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, X. R., Chang-Claude, J., Goode, E. L., Couch, F. J., Nevanlinna, H., Milne, R. L., ... Garcia-Closas, M. (2011). Associations of breast cancer risk factors with tumor subtypes: A pooled analysis from the Breast Cancer Association Consortium studies. Journal of the National Cancer Institute, 103(3), 250–263. https://doi.org/10.1093/jnci/djx526
Yovel, Y., Franz, M. O., Stilz, P., & Schnitzler, H.-U. (2008). Plant classification from bat-like echolocation signals. *PLOS Computational Biology, 4*(3):e1000032. https://doi.org/10.1371/journal.pcbi.1000032

**SUPPORTING INFORMATION**
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Feng H, Gusev A, Pasaniuc B, et al. Transcriptome-wide association study of breast cancer risk by estrogen-receptor status. *Genetic Epidemiology*. 2020;44:442–468. https://doi.org/10.1002/gepi.22288