UCLA
UCLA Previously Published Works

Title

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

Permalink
https://escholarship.org/uc/item/3v22d5nx

Journal
Genetic Epidemiology, 44(5)

ISSN
0741-0395

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

Publication Date
2020-07-01

DOI
10.1002/gepi.22288

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

HHS Public Access

Author manuscript

Genet Epidemiol. Author manuscript; available in PMC 2021 March 23.

Published in final edited form as:

Genet Epidemiol. 2020 July ; 44(5): 442–468. doi:10.1002/gepi.22288.

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

A full list of authors and affiliations appears at the end of the article.

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

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.

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

SUPPORTING INFORMATION

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 straight-forward (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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 Haplotype 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 .8), minor allele frequency (MAF) 0.05, and were included in the Hap-Map 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 $r^2 > .3$, and an 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 *cis* (500 base-pair window surrounding transcription start site) SNP-heritability $\left(cis - h_g^2\right)$ for each transcript expression (Cai et al., 2014; J. Yang et al., 2010). Only genes with significant heritability (nominal p .01) were included in the subsequent model construction (J. Yang, Lee, Goddard, & Visscher, 2011). The p values for null hypotheses $cis - h_g^2 = 0$ 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 $cis - h_g^2$ 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 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, $\Sigma_{s,s}$ be the LD matrix from reference genotype data and let $W = (w_1 w_2 w_3 \ldots w_j)$ 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 0 is associated with disease, the test statistic $wz/(w\Sigma_{s,s}w')^{1/2}$ follows a normal distribution with mean = 0 and variance = 1. To account for finite sample size and instances where $\Sigma_{s,s}$ was not invertible, we adjusted the diagonal of the matrix using a technique similar to ridge regression with $\lambda = 0.1$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

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 β_2 for ER− disease and β_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 $cis - h_g^2 < .01$). We then built linear predictors for these heritable genes and estimated prediction r^2 using fivefold cross-validation. A total of 454 genes failed our cross-validation r^2 requirement ($r^2 > .01$), and we performed TWAS on the remaining 901 genes. We defined statistical significance for TWAS results as a marginal $p < 5.5 \times 10^{-5}$ (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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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-tringent 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 ($p < .05/901 = 5.5 \times 10^{-5}$; 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 $r^2 > .1$; 383 genes passed this threshold), we found four TWAS significant ($p < .05/383 = 1.3 \times 10^{-4}$) 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 GWAS SNP remained significantly associated with breast cancer risk at the genome-wide threshold (5×10^{-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 (*RP11-73O6.3*, *L3MBTL3*),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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, *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).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 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 3.70×10^{-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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Authors

Helian Feng[1,2,3], Alexander Gusev[4], Bogdan Pasaniuc[5], Lang Wu[6], Jirong Long[7], Zomoroda Abu-full[8], Kristiina Aittomäki[9], Irene L. Andrulis[10,11], Hoda Anton-Culver[12], Antonis C. Antoniou[13], Adalgeir Arason[14,15], Volker Arndt[16], Kristan J. Aronson[17], Banu K. Arun[18], Ella Asseryanis[19], Paul L. Auer[20,21], Jacopo Azzollini[22], Judith Balmaña[23], Rosa B. Barkardottir[14,15], Daniel R. Barnes[13], Daniel Barrowdale[13], Matthias W. Beckmann[24], Sabine Behrens[25], Javier Benitez[26,27], Marina Bermisheva[28], Katarzyna Białkowska[29], Ana Blanco[26,30,31], Carl Blomqvist[32,33], Bram Boeckx[34,35], Natalia V. Bogdanova[36,37,38], Stig E. Bojesen[39,40,41], Manjeet K. Bolla[13], Bernardo Bonanni[42], Ake Borg[43], Hiltrud Brauch[44,45,46], Hermann Brenner[16,46,47], Ignacio Briceno[48,49], Annegien Broeks[50], Thomas Brüning[51], Barbara Burwinkel[52,53], Qiuyin Cai[7], Trinidad Caldés[54], Maria A. Caligo[55], Ian Campbell[56,57], Sander Canisius[50,58], Daniele Campa[59], Brian D. Carter[60], Jonathan Carter[61], Jose E. Castelao[62], Jenny Chang-Claude[25,63], Stephen J. Chanock[64], Hans Christiansen[36], Wendy K. Chung[65], Kathleen B. M. Claes[66], Christine L. Clarke[67], GEMO Study Collaborators[68,69,70], EMBRACE Collaborators[13], GC-HBOC study Collaborators[71], Fergus J. Couch[72], Angela Cox[73], Simon S. Cross[74], Cezary Cybulski[29], Kamila Czene[75], Mary B. Daly[76], Miguel de la Hoya[54], Kim De Leeneer[66], Joe Dennis[13], Peter Devilee[77,78], Orland Diez[79,80], Susan M. Domchek[81], Thilo Dörk[37], Isabel dos-Santos-Silva[82], Alison M. Dunning[83], Miriam Dwek[84], Diana M. Eccles[85], Bent Ejlertsen[86], Carolina Ellberg[87], Christoph Engel[88,89], Mikael Eriksson[75], Peter A. Fasching[24,90], Olivia Fletcher[91], Henrik Flyger[92], Florentia Fostira[93], Eitan Friedman[94,95], Lin Fritschi[96], Debra Frost[13], Marike Gabrielson[75], Patricia A. Ganz[97], Susan M. Gapstur[60], Judy Garber[98], Montserrat García-Closas[64,99,100], José A. García-Sáenz[54], Mia M. Gaudet[60], Graham G. Giles[101,102], Gord Glendon[10], Andrew K. Godwin[103], Mark S. Goldberg[104,105], David E. Goldgar[106], Anna González-Neira[27], Mark H. Greene[107], Jacek Gronwald[29], Pascal Guénel[108], Christopher A. Haiman[109], Per Hall[75,110], Ute Hamann[111], Christopher Hake[112], Wei He[75], Jane Heyworth[113], Frans B.L. Hogervorst[114], Antoinette Hollestelle[115], Maartje J. Hooning[115], Robert N. Hoover[64], John L. Hopper[101], Guanmengqian Huang[111], Peter J. Hulick[116,117], Keith Humphreys[75], Evgeny N. Imyanitov[118], ABCTB Investigators[119], HEBON Investigators[120], BCFR Investigators[121], OCGN Investigators[122], Claudine Isaacs[123], Milena Jakimovska[124], Anna Jakubowska[29,125], Paul James[57,126], Ramunas Janavicius[127], Rachel C. Jankowitz[128], Esther M. John[121], Nichola Johnson[91], Vijai Joseph[129], Audrey Jung[25], Beth Y. Karlan[130], Elza Khusnutdinova[28,131], Johanna I. Kiiski[132], Irene Konstantopoulou[93], Vessela N. Kristensen[133,134], Yael Laitman[94], Diether Lambrechts[34,35], Conxi Lazaro[135], Dominique Leroux[136], Goska Leslie[13], Jenny Lester[130], Fabienne Lesueur[69,70,137], Noralane Lindor[138], Sara Lindström[139,140], Wing-Yee Lo[44,45], Jennifer T. Loud[107], Jan Lubi ski[29], Enes Makalic[101], Arto Mannermaa[141,142,143], Mehdi Manoochehri[111], Siranoush Manoukian[22], Sara Margolin[110,144], John W.M. Martens[115], Maria E. Martinez[145,146], Laura Matricardi[147], Tabea Maurer[63], Dimitrios Mavroudis[148], Lesley McGuffog[13], Alfons Meindl[149], Usha Menon[150],

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Kyriaki Michailidou[13,151], Pooja M. Kapoor[25,152], Austin Miller[153], Marco Montagna[147], Fernando Moreno[54], Lidia Moserle[147], Anna M. Mulligan[154,155], Taru A. Muranen[132], Katherine L. Nathanson[81], Susan L. Neuhausen[156], Heli Nevanlinna[132], Ines Nevelsteen[157], Finn C. Nielsen[158], Liene Nikitina-Zake[159], Kenneth Offit[129,160], Edith Olah[161], Olufunmilayo I. Olopade[162], Håkan Olsson[87], Ana Osorio[26,27], Janos Papp[161], Tjoung-Won Park-Simon[37], Michael T. Parsons[163], Inge S. Pedersen[164], Ana Peixoto[165], Paolo Peterlongo[166], Julian Peto[82], Paul D.P. Pharoah[13,83], Kelly-Anne Phillips[56,57,101,167], Dijana Plaseska-Karanfilska[124], Bruce Poppe[66], Nisha Pradhan[129], Karolina Prajzendanc[29], Nadege Presneau[84], Kevin Punie[157], Katri Pylkäs[168,169], Paolo Radice[170], Johanna Rantala[171], Muhammad Usman Rashid[111,172], Gad Rennert[8], Harvey A Risch[173], Mark Robson[160], Atocha Romero[174], Emmanouil Saloustros[175], Dale P. Sandler[176], Catarina Santos[165], Elinor J. Sawyer[177], Marjanka K. Schmidt[50,178], Daniel F. Schmidt[101,179], Rita K. Schmutzler[71,180], Minouk J. Schoemaker[99], Rodney J Scott[181,182,183], Priyanka Sharma[184], Xiao-Ou Shu[7], Jacques Simard[185], Christian F. Singer[19], Anne-Bine Skytte[186], Penny Soucy[185], Melissa C. Southey[187,188], John J. Spinelli[189,190], Amanda B. Spurdle[163], Jennifer Stone[101,191], Anthony J. Swerdlow[99,192], William J. Tapper[193], Jack A. Taylor[176,194], Manuel R. Teixeira[165,195], Mary Beth Terry[196], Alex Teulé[197], Mads Thomassen[198], Kathrin Thöne[63], Darcy L. Thull[199], Marc Tischkowitz[200,201], Amanda E. Toland[202], Rob A. E. M. Tollenaar[203], Diana Torres[48,111], Thérèse Truong[108], Nadine Tung[204], Celine M. Vachon[205], Christi J. van Asperen[206], Ans M. W. van den Ouweland[207], Elizabeth J. van Rensburg[208], Ana Vega[26,30,31], Alessandra Viel[209], Paula Vieiro-Balo[210], Qin Wang[13], Barbara Wappenschmidt[71,180], Clarice R. Weinberg[211], Jeffrey N. Weitzel[212], Camilla Wendt[144], Robert Winqvist[168,169], Xiaohong R. Yang[64], Drakoulis Yannoukakos[93], Argyrios Ziogas[12], Roger L. Milne[100,101,187], Douglas F. Easton[13,83], Georgia Chenevix-Trench[163], Wei Zheng[7], Peter Kraft[1,2], Xia Jiang[1,2]

Affiliations

[1]Program in Genetic Epidemiology and Statistical Genetics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts [2]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts [3]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts [4]Dana-Farber Cancer Institute, Boston, Massachusetts [5]UCLA Path & Lab Med, Los Angeles, California [6]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii [7]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee [8]Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel [9]Department of Clinical Genetics, Helsinki University Hospital, University of Helsinki, Helsinki, Finland [10]Fred A, Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada [11]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada [12]Department of Epidemiology, Genetic Epidemiology Research Institute, University

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of California Irvine, Irvine, California [13]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [14]Department of Pathology, Landspitali University Hospital, Reykjavik, Iceland [15]BMC (Biomedical Centre), Faculty of Medicine, University of Iceland, Reykjavik, Iceland [16]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany [17]Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, Ontario, Canada [18]Department of Breast Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas [19]Department of OB/GYN and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria [20]Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington [21]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin [22]Unit of Medical Genetics, Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy [23]High Risk and Cancer Prevention Group, Vall d'Hebron Institute of Oncology, Barcelona, Spain [24]Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany [25]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany [26]Centro de Investigaci—n en Red de Enfermedades Raras (CIBERER), Madrid, Spain [27]Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain [28]Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia [29]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland [30]Fundaci—n Pœblica Galega Medicina Xen—mica, Santiago De Compostela, Spain [31]Instituto de Investigacion Sanitaria de Santiago de Compostela, Santiago de Compostela, Spain [32]Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland [33]Department of Oncology, University Hospital, Karolinska Institute, Stockholm, Sweden [34]VIB Center for Cancer Biology, VIB, Leuven, Belgium [35]Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium [36]Department of Radiation Oncology, Hannover Medical School, Hannover, Germany [37]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany [38]NN Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus [39]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark [40]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark [41]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark [42]Division of Cancer Prevention and Genetics, IEO, European Institute of Oncology IRCCS, Milan, Italy [43]Department of Oncology, Lund University and Skåne University Hospital, Lund, Sweden [44]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany [45]iFIT-Cluster of Excellence, University of Tuebingen, Tuebingen, Germany [46]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany [47]Division of Preventive

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany [48]Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota, Colombia [49]Medical Faculty, Universidad de La Sabana, Bogota, Colombia [50]Division of Molecular Pathology, The Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands [51]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [52]Molecular Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany [53]Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Germany [54]Medical Oncology Department, Hospital Cl'nico San Carlos, Instituto de Investigaci—n Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain [55]Section of Molecular Genetics, Dept, of Laboratory Medicine, University Hospital of Pisa, Pisa, Italy [56]Research Department, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia [57]Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Victoria, Australia [58]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute —Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands [59]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany [60]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, Georgia [61]Department of Gynaecological Oncology, Chris OÕBrien Lifehouse and The University of Sydney, Camperdown, New South Wales, Australia [62]Oncology and Genetics Unit, Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain [63]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany [64]Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, Maryland [65]Departments of Pediatrics and Medicine, Columbia University, New York, New York [66]Centre for Medical Genetics, Ghent University, Gent, Belgium [67]Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia [68]Department of Tumour Biology, INSERM U830, Paris, France [69]Institut Curie, Paris, France [70]Mines ParisTech, Fontainebleau, France [71]Center for Hereditary Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany [72]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota [73]Department of Oncology and Metabolism, Sheffield Institute for Nucleic Acids (SInFoNiA), University of Sheffield, Sheffield, UK [74]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK [75]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden [76]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, Pennsylvania [77]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands [78]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands [79]Hereditary Cancer Genetics Group, Area of Clinical and Molecular Genetics, Vall dHebron Institute of Oncology

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

(VHIO), University Hospital Vall d'Hebron, Barcelona, Spain [80]Clinical and Molecular Genetics Area, University Hospital Vall dHebron, Barcelona, Spain [81]Department of Medicine, Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania [82]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK [83]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK [84]Department of Biomedical Sciences, Faculty of Science and Technology, University of Westminster, London, UK [85]Cancer Sciences Academic Unit, Faculty of Medicine, University of Southampton, Southampton, UK [86]Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark [87]Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden [88]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany [89]LIFE - Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany [90]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, California [91]The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK [92]Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark [93]Molecular Diagnostics Laboratory, INRASTES, National Centre for Scientific Research 'Demokritos', Athens, Greece [94]The Susanne Levy Gertner Oncogenetics Unit, Chaim Sheba Medical Center, Ramat Gan, Israel [95]Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel [96]School of Public Health, Curtin University, Perth, Western Australia, Australia [97]Schools of Medicine and Public Health, Division of Cancer Prevention & Control Research, Jonsson Comprehensive Cancer Centre, UCLA, Los Angeles, California [98]Cancer Risk and Prevention Clinic, Dana-Farber Cancer Institute, Boston, Massachusetts [99]Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK [100]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia [101]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia [102]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia [103]Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Kansas City, Kanas [104]Department of Medicine, McGill University, Montreal, Quebec, Canada [105]Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montreal, Quebec, Canada [106]Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah [107]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland [108]Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France [109]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California [110]Department of Oncology, Sšdersjukhuset, Stockholm, Sweden [111]Molecular Genetics of Breast Cancer,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

German Cancer Research Center (DKFZ), Heidelberg, Germany [112]City of Hope Clinical Cancer Genetics Community Research Network, Duarte, California [113]School of Population and Global Health, The University of Western Australia, Perth, Western Australia, Australia [114]Family Cancer Clinic, The Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands [115]Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands [116]Center for Medical Genetics, NorthShore University HealthSystem, Evanston, Illinois [117]The University of Chicago Pritzker School of Medicine, Chicago, Illinois [118]NN Petrov Institute of Oncology, St. Petersburg, Russia [119]Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia [120]The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), Coordinating Center, The Netherlands Cancer Institute, Amsterdam, The Netherlands [121]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California [122]Ontario Cancer Genetics Network, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada [123]Lombardi Comprehensive Cancer Center, Georgetown University, Washington, District of Columbia [124]Research Centre for Genetic Engineering and Biotechnology 'Georgi D, Efremov', Macedonian Academy of Sciences and Arts, Skopje Republic of North Macedonia, North Macedonia [125]Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland [126]Parkville Familial Cancer Centre, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia [127]State Research Institute Innovative Medicine Center, Vilnius, Lithuania [128]Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania [129]Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York [130]David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California, Los Angeles, California [131]Department of Genetics and Fundamental Medicine, Bashkir State Medical University, Ufa, Russia [132]Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland [133]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway [134]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway [135]Molecular Diagnostic Unit, Hereditary Cancer Program, ICO-IDIBELL (Bellvitge Biomedical Research Institute, Catalan Institute of Oncology), CIBERONC, Barcelona, Spain [136]Departement de Ge netique, CHU de Grenoble, Grenoble, France [137]Genetic Epidemiology of Cancer Team, Inserm U900, Paris, France [138]Department of Health Sciences Research, Mayo Clinic, Scottsdale, Arizona [139]Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington [140]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington [141]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland [142]Institute of Clinical Medicine, Pathology and Forensic Medicine,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

University of Eastern Finland, Kuopio, Finland [143]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland [144]Department of Clinical Science and Education, Sšdersjukhuset, Karolinska Institutet, Stockholm, Sweden [145]Moores Cancer Center, University of California San Diego, La Jolla, California [146]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California [147]Immunology and Molecular Oncology Unit, Veneto Institute of Oncology ÊIOV—IRCCS, Padua, Italy [148]Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece [149]Department of Gynecology and Obstetrics, Ludwig Maximilian University of Munich, Munich, Germany [150]MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, University College London, London, UK [151]Department of Electron Microscopy/Molecular Pathology and The Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus [152]Faculty of Medicine, University of Heidelberg, Heidelberg, Germany [153]NRG Oncology, Statistics and Data Management Center, Roswell Park Cancer Institute, Buffalo, New York [154]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada [155]Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada [156]Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, California [157]Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium [158]Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark [159]Latvian Biomedical Research and Study Centre, Riga, Latvia [160]Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York [161]Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary [162]Center for Clinical Cancer Genetics, The University of Chicago, Chicago, Illinois [163]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia [164]Section of Molecular Diagnostics, Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark [165]Department of Genetics, Portuguese Oncology Institute, Porto, Portugal [166]Genome Diagnostics Program, IFOM—The FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy [167]Department of Medicine Oncology, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia [168]Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland [169]Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland [170]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy [171]Clinical Genetics, Karolinska Institutet, Stockholm, Sweden [172]Department of Basic Sciences, Shaukat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Lahore, Pakistan [173]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut [174]Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain [175]Department of Oncology, University Hospital of Larissa, Larissa, Greece

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[176]Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina [177]Research Oncology, GuyÕs Hospital, King's College London, London, UK [178]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands [179]Faculty of Information Technology, Monash University, Melbourne, Victoria, Australia [180]Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany [181]Division of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, New South Wales, Australia [182]Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Callaghan, New South Wales, Australia [183]Hunter Medical Research Institute, John Hunter Hospital, Newcastle, New South Wales, Australia [184]Department of Internal Medicine, Division of Medical Oncology, University of Kansas Medical Center, Westwood, Kanas [185]Genomics Center, Centre Hospitalier Universitaire de Quebec–Universite Laval, Research Center, Quebec City, Qubec, Canada [186]Department of Clinical Genetics, Aarhus University Hospital, Aarhus N, Denmark [187]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia [188]Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia [189]Population Oncology, BC Cancer, Vancouver, British of Columbia, Canada [190]School of Population and Public Health, University of British Columbia, Vancouver, British of Columbia, Canada [191]The Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and University of Western Australia, Perth, Western Australia, Australia [192]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK [193]Faculty of Medicine, University of Southampton, Southampton, UK [194]Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina [195]Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal [196]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York [197]Genetic Counseling Unit, Hereditary Cancer Program, IDIBELL (Bellvitge Biomedical Research Institute), Catalan Institute of Oncology, CIBERONC, Barcelona, Spain [198]Department of Clinical Genetics, Odense University Hospital, Odence C, Denmark [199]Department of Medicine, Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania [200]Program in Cancer Genetics, Departments of Human Genetics and Oncology, McGill University, Montreal, Quebec, Canada [201]Department of Medical Genetics, University of Cambridge, Cambridge, UK [202]Department of Cancer Biology and Genetics, The Ohio State University, Columbus, Ohio [203]Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands [204]Department of Medical Oncology, Beth Israel Deaconess Medical Center, Boston, Massachusetts [205]Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, Minnesota [206]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands [207]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The Netherlands [208]Department of Genetics, University of Pretoria, Arcadia, South Africa [209]Division of Functional Onco-genomics and Genetics, Centro di Riferimento Oncologico di Aviano (CRO), IRCCS, Aviano, Italy [210]Hospital Clínico Universitario (SERGAS), Universidad de Santiago de Compostela, CIMUS, Santiago de Compostela, España [211]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina [212]Clinical Cancer Genomics, City of Hope, Duarte, California

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 PSRSIIRI-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 (1U19 CA148537, 1U19 CA148065, and 1U19 CA148112—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 (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 ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J. L. H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M. C. S. is an NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society (grants NKI 2007-3839; 2009 4363). The Australian Breast Cancer Tissue Bank (ABCTB) is generously supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The ACP study is funded by the Breast Cancer Research Trust, UK. The AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute (grant no. Z01-CP010119), and the National Institute of Environmental Health Sciences (grant no. Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BBCS 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 BCEES 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 UM1 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 GA-lician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER;

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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, Consellería 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 Economia 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'Education 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 Deutsche Krebshilfe e. V. (70492) and the German Cancer Research Center (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 Juselius Foundation. The HERPACC was supported by MEXT Kakenhi (No. 170150181 and 26253041) from the Ministry of Education, Science, Sports, Culture and Technology of Japan, by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from Ministry Health, Labour and Welfare of Japan, by Health and Labour Sciences Research Grants for Research on Applying Health Technology from Ministry Health, Labour and Welfare of Japan, by National Cancer Center Research and Development Fund, and "Practical Research for Innovative Cancer Control (15ck0106177h0001)" from Japan Agency for Medical Research and development, AMED, and Cancer Bio Bank Aichi. The HMBCS was supported by a grant from the Friends of Hannover Medical School and by the Rudolf Bartling Foundation. The HUBCS was supported by a grant from the German Federal Ministry of Research and Education (RUS08/017) and by the Russian Foundation for Basic Research and the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Federal Agency for Scientific Organizations for supporting the Bioresource collections and RFBR grants 14-04-97088, 17-29-06014, and 17-44-020498. ICICLE 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. Financial support for KAR-BAC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and Bert von Kantzows foundation. The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and the strategic funding of the University of Eastern Finland.

The kConFab Follow-Up Study is supported by grants from Cancer Australia, the Australian National Breast Cancer Foundation, the National Health and Medical Research Council, the National Institute of Health USA, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. KAP is an Australian National Breast Cancer Foundation Practitioner Fellow. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command (DAMD17-01-1-0729), Cancer Council Victoria, Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, and 199600). G. C. T. and P. W. are supported by the NHMRC. R. B. was a Cancer Institute NSW Clinical Research Fellow. 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 I, 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-SIIRI-701. MYBRCA is funded by research grants from the Malaysian Ministry of Higher Education (UM.C/HlR/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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 (RUL 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, 2013PR044. PROCAS is funded from NIHR grant PGfAR 0707-10031. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The SASBAC 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 by the Ministry of Education, Science and Technology (2012-0000347). SGBCC is funded by the NUS start-up Grant, National University Cancer Institute Singapore (NCIS) Centre Grant and the NMRC Clinician Scientist Award. Additional controls were recruited by the Singapore Consortium of Cohort Studies-Multi-ethnic cohort (SCCS-MEC), which was funded by the Biomedical Research Council, grant no. 05/1/21/19/425. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The Two Sister Study (2SISTER) was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES102245), and, also by a grant from Susan G. Komen for the Cure, grant FAS0703856. SKKDKFZS is supported by the DKFZ. The SMC is funded by the Swedish Cancer Foundation. The SZBCS and IHCC were supported by Grant PBZ_KBN_122/P05/2004 and the program of the Minister of Science and Higher Education under the name “Regional Initiative of Excellence” in 2019-2022 project number 002/RID/2018/19 amount of financing 12,000,000 PLN. The TBCS was funded by The National Cancer Institute of Thailand. The TNBCC was supported by a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), a grant from the Breast Cancer Research Foundation, a generous gift from the David F. and Margaret T. Grohne Family Foundation. The TWBCS is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences, Academia Sinica, Taiwan. The UCIBCS component of this research was supported by the NIH (CA58860 and CA92044) and the Lon V Smith Foundation (LVS39420). The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London and also thank the study participants, study staff, and the doctors, nurses and other health care providers 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 USRT 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 (HHSN268201100046C, HHSN268201 100001C, 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-1215-20007). 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 Shared Resource at Georgetown University (NIH/NCI P30-CA51008) 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" (PI10/01422, PI13/00285, 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 Enferemdades Raras CIBERER (ACCI 2016: ER17P1AC7112/2018); 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.

OSUCCC was funded by the Ohio State University Comprehensive Cancer Center. Leigha Senter, Kevin Sweet, Caroline Craven, Julia Cooper, Amber Aielts, 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, N02-CP-21013-63 and N02-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. Collée, 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. MeijersHeijboer, T.A.M. van Os; VU University Medical Center, Amsterdam, NL: K. van Engelen, J.J.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, NKI20073756, 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.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/ (Michailidou et al., 2017, pp. 92–94).

REFERENCES

Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, … Easton DF (2016). The OncoArray Consortium: A network for understanding the genetic architecture of common cancers. Cancer Epidemiology, Biomarkers & Prevention, 10.1158/1055-9965.EPI-16-0106

Antoniou AC, Beesley J, McGuffog L, Sinilnikova OM, Healey S, Neuhausen SL, … 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. 10.1158/0008-5472.CAN-10-1907 [PubMed: 21118973]

Antoniou AC, Goldgar DE, Andrieu N, Chang-Claude J, Brohet R, Rookus MA, & Easton DF (2005). A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. Genetic Epidemiology, 29(1), 1–11. 10.1002/gepi.20074 [PubMed: 15880399]

Atchley DP, Albarracin CT, Lopez A, Valero V, Amos CI, Gonzalez-Angulo AM, … Arun BK (2008). Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. Journal of Clinical Oncology, 26(26), 4282–4288. 10.1200/JCO.2008.16.6231 [PubMed: 18779615]

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. 10.1002/gepi.22131 [PubMed: 29808603]

Barnes DR, Lee A, EMBRACE Investigators, kConFab Investigators, Easton DF, & Antoniou AC (2012). Evaluation of association methods for analysing modifiers of disease risk in carriers of high-risk mutations. Genetic Epidemiology, 36(3), 274–291. 10.1002/gepi.21620 [PubMed: 22714938]

Beggs AD, & Hodgson SV (2009). Genomics and breast cancer: The different levels of inherited susceptibility. European Journal of Human Genetics, 17(7), 855–856. 10.1038/ejhg.2008.235 [PubMed: 19092772]

Bertolini F (2013). Adipose tissue and breast cancer progression: A link between metabolism and cancer. Breast, 22(Suppl 2), S48–S49. 10.1016/j.breast.2013.07.009 [PubMed: 24074792]

Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, 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. 10.1371/journal.pmed.1000279 [PubMed: 20520800]

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. 10.1186/1897-4287-11-12 [PubMed: 24025454]

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, & 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. 10.3322/caac.21492 [PubMed: 30207593]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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. 10.1038/ng.3041 [PubMed: 25038754]

Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MWR, Pooley KA, … 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. 10.1038/ng1981 [PubMed: 17293864]

Darabi H, Beesley J, Droit A, Kar S, Nord S, Moradi Marjaneh M, … Dunning AM (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. 10.1038/srep32512 [PubMed: 27600471]

Delaneau O, Marchini J, & Zagury J-F (2011). A linear complexity phasing method for thousands of genomes. Nature Methods, 9(2), 179–181. 10.1038/nmeth.1785 [PubMed: 22138821]

Dunning AM, Michailidou K, Kuchenbaecker KB, Thompson D, French JD, Beesley J, … Edwards SL (2016). Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nature Genetics, 48(4), 374–386. 10.1038/ng.3521 [PubMed: 26928228]

Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, … Ponder BAJ (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature, 447(7148), 1087–1093. 10.1038/nature05887 [PubMed: 17529967]

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57–74. 10.1038/nature11247 [PubMed: 22955616]

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 MA, Gamazon ER, Al-Ejeh F, Aittomäki K, Andrulis IL, Anton-Culver H, … Chenevix-Trench G (2019). Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. Nature Communications, 10(1), 1741. 10.1038/s41467-018-08053-5

Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, … Im HK (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics, 47(9), 1091–1098. 10.1038/ng.3367 [PubMed: 26258848]

Gao G, Pierce BL, Olopade OI, Im HK, & 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. 10.1371/journal.pgen.1006727 [PubMed: 28957356]

Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, … Kraft P (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nature Genetics, 45(4), 392–398, 398e1–2. 10.1038/ng.2561 [PubMed: 23535733]

de Graauw M, Cao L, Winkel L, van Miltenburg M. H. a M., le Dévédec SE, Klop M, … van de Water B (2014). Annexin A2 depletion delays EGFR endocytic trafficking via cofilin activation and enhances EGFR signaling and metastasis formation. Oncogene, 33(20), 2610–2619. 10.1038/onc.2013.219 [PubMed: 23792445]

GTEx Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348(6235), 648–660. 10.1126/science.1262110 [PubMed: 25954001]

Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, … Pasaniuc B (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics, 48(3), 245–252. 10.1038/ng.3506 [PubMed: 26854917]

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, … Young RA (2013). Super-enhancers in the control of cell identity and disease. Cell, 155(4), 934–947. 10.1016/j.cell.2013.09.053 [PubMed: 24119843]

Hoffman JD, Graff RE, Emami NC, Tai CG, Passarelli MN, Hu D, … Witte JS (2017). Cis-eQTL-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk. PLOS Genetics, 13(3):e1006690. 10.1371/journal.pgen.1006690 [PubMed: 28362817]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Howie BN, Donnelly P, & Marchini J (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLOS Genetics, 5(6): e1000529. 10.1371/journal.pgen.1000529 [PubMed: 19543373]

Jakubowska A, Gronwald J, Menkiszak J, Górski B, Huzarski T, Byrski T, … Hamann U (2010). BRCA1associated breast and ovarian cancer risks in Poland: No association with commonly studied polymorphisms. Breast Cancer Research and Treatment, 119(1), 201–211. 10.1007/s10549-009-0390-5 [PubMed: 19360465]

Lin W-Y, Camp NJ, Ghoussaini M, Beesley J, Michailidou K, Hopper JL, … Cox A (2015). Identification and characterization of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. Human Molecular Genetics, 24(1), 285–298. 10.1093/hmg/ddu431 [PubMed: 25168388]

Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, & Pasaniuc B (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nature Genetics, 51(4), 675–682. 10.1038/s41588-019-0367-1 [PubMed: 30926970]

Mancuso 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. 10.1016/j.ajhg.2017.01.031 [PubMed: 28238358]

Mavaddat N, Antoniou AC, Easton DF, & Garcia-Closas M (2010). Genetic susceptibility to breast cancer. Molecular oncology, 4(3), 174–191. 10.1016/j.molonc.2010.04.011 [PubMed: 20542480]

Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, … Garcia-Closas M (2015). Prediction of breast cancer risk based on profiling with common genetic variants. Journal of the National Cancer Institute, 107(5), 10.1093/jnci/djv036

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, … Mahajan A, Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics, 48(10), 1279–1283. 10.1038/ng.3643 [PubMed: 27548312]

Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, … Easton DF (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. 10.1038/ng.3242 [PubMed: 25751625]

Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, … Easton DF (2017). Association analysis identifies 65 new breast cancer risk loci. Nature, 551(7678), 92–94. 10.1038/nature24284 [PubMed: 29059683]

Milne RL, Kuchenbaecker KB, 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. 10.1038/ng.3785 [PubMed: 29058716]

Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, & Cox NJ (2010). Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. PLOS Genetics, 6(4):e1000888. 10.1371/journal.pgen.1000888 [PubMed: 20369019]

Pasaniuc B, Zaitlen N, Shi H, Bhatia G, Gusev A, Pickrell J, … Price AL (2014). Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics, 30(20), 2906–2914. 10.1093/bioinformatics/btu416 [PubMed: 24990607]

Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, … Rahman N (2006). Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nature Genetics, 38(11), 1239–1241. 10.1038/ng1902 [PubMed: 17033622]

Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, … Easton DF (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. Nature Genetics, 42(6), 504–507. 10.1038/ng.586 [PubMed: 20453838]

Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, … Kundaje A (2019). Opportunities and challenges for transcriptome-wide association studies. Nature Genetics, 51(4), 592–599. 10.1038/s41588-019-0385-z [PubMed: 30926968]

Warren Andersen S, Trentham-Dietz A, Gangnon RE, Hampton JM, Figueroa JD, Skinner HG, … Newcomb PA (2013). The associations between a polygenic score, reproductive and menstrual risk

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

factors and breast cancer risk. Breast Cancer Research and Treatment, 140(2), 427–434. 10.1007/s10549-013-2646-3 [PubMed: 23893088]

Willer CJ, Li Y, & Abecasis GR (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26(17), 2190–2191. 10.1093/bioinformatics/btq340 [PubMed: 20616382]

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. 10.1038/s41588-018-0132-x [PubMed: 29915430]

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, … Visscher PM (2010). Common SNPs explain a large proportion of the heritability for human height. Nature Genetics, 42(7), 565–569. 10.1038/ng.608 [PubMed: 20562875]

Yang J, Lee SH, Goddard ME, & Visscher PM (2011). GCTA: A tool for genome-wide complex trait analysis. American Journal of Human Genetics, 88(1), 76–82. 10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]

Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, … 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. 10.1093/jnci/djq526 [PubMed: 21191117]

Yovel Y, Franz MO, Stilz P, & Schnitzler H-U (2008). Plant classification from bat-like echolocation signals. PLOS Computational Biology, 4(3):e1000032. 10.1371/journal.pcbi.1000032 [PubMed: 18369425]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

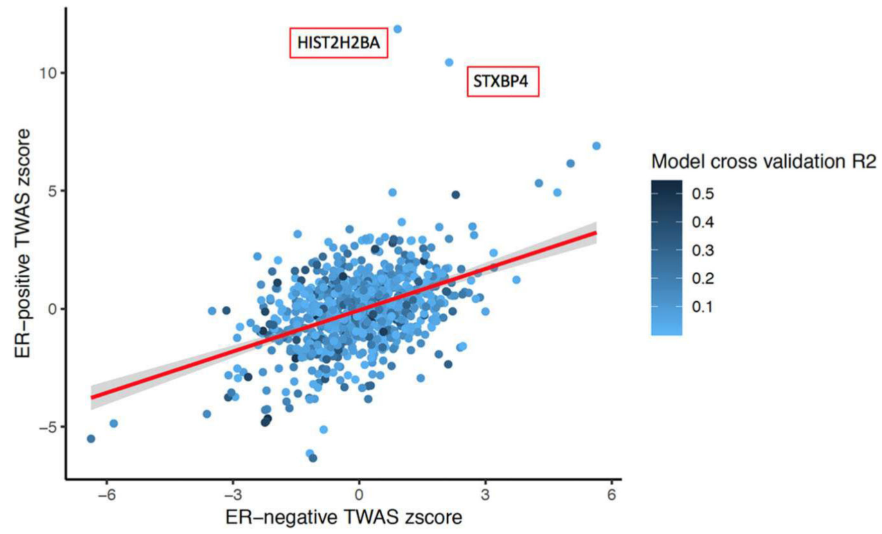

FIGURE 1.

Scatter plot comparing the transcriptome-wide association study *z* scores in ER+ and ER− patients

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

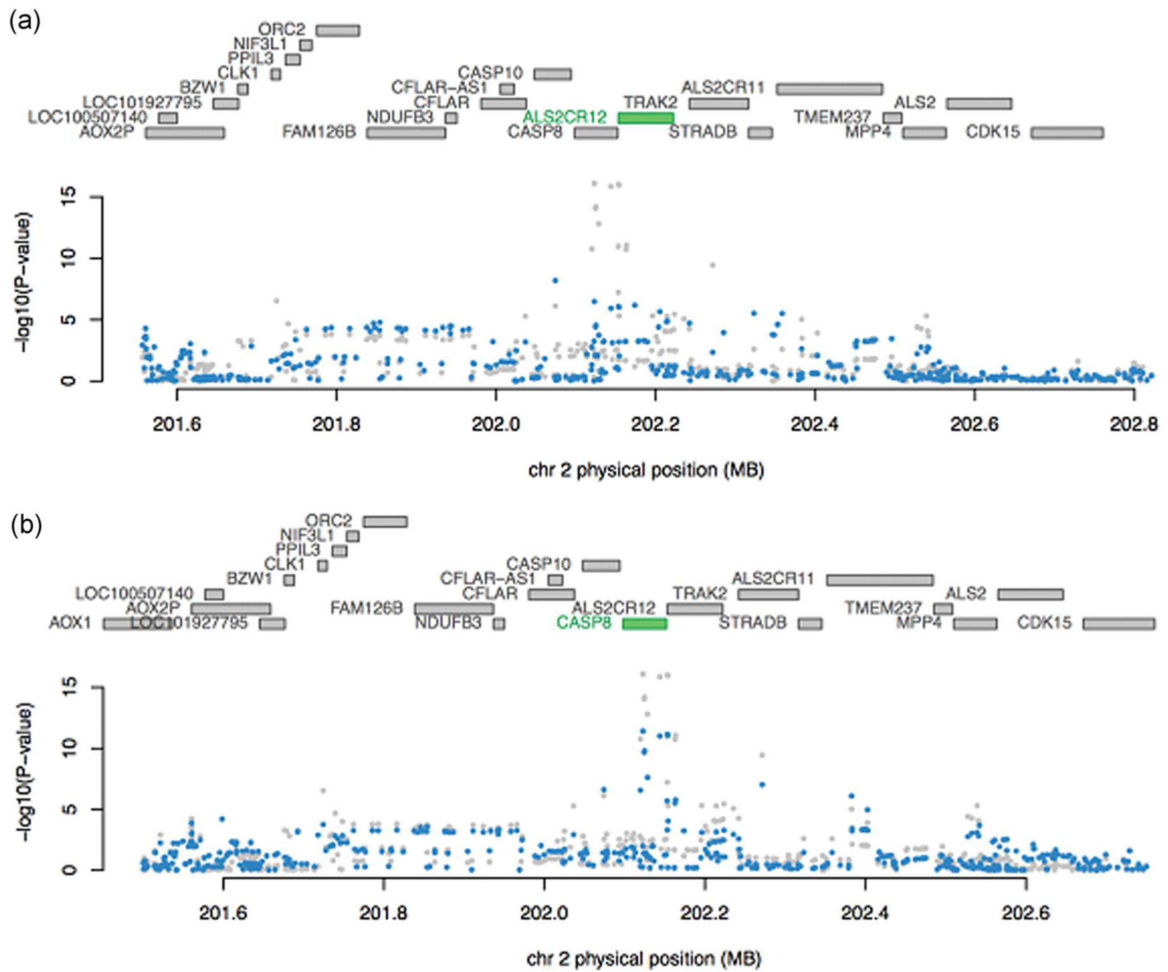

FIGURE 2.

Conditional and joint analysis (COJO) for genes near a strong breast cancer GWAS hit. (a) COJO results adjusting for predicted expression of *ALS2CR12*. After conditioning on *ALS2CR12*, almost all original significant GWAS signals (grey dots) disappear (blue dots). (b) COJO results adjusting for the predicted expression of *CASP8*. After conditioning on *CASP8*, some of the original GWAS significant signals (grey dots) remains (blue dots)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

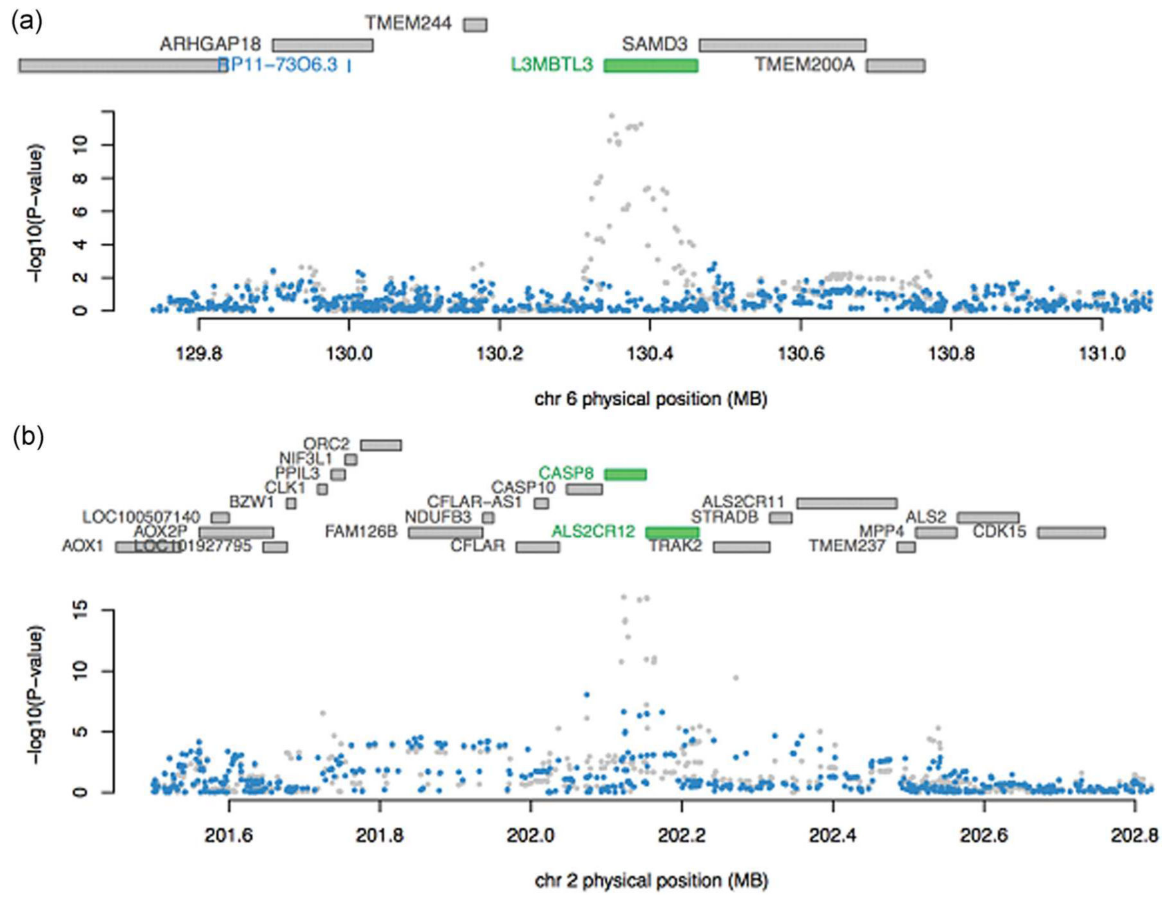

FIGURE 3.

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 *ALS2C12*). 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 remain significant after COJO). COJO, conditional and joint analysis; GWAS, genome-wide association studies; TWAS, transcriptome-wide association studies

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 1

Genes significantly associated with the overall breast cancer, estrogen receptor positive and negative subtypes, and estrogen receptor status, as identified by TWAS

						TWAS *p* values			
Cytoband	**Gene**	**Chromosome: Position (start–end)**	**Number of SNPs**	**Heritability**	**Cross validation r^2**	**Overall vs. controls**	**ER+ vs. controls**	**ER− vs. controls**	**ER+ vs. ER−**
p11.2	HIST2H2BA	1: 120906028–120915073	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	RP11–73O6.3	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: 5807186–5884095	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-2323K18.1	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: 44370099–44415160	75	0.27	0.20	**7.9E–08**	**2.5E–05**	9.7E–02	4.5E–01
	LRRC37A2	17: 44588877–44630815	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**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cytoband	Gene	Chromosome: Position (start–end)	Number of SNPs	Heritability	Cross validation r^2	TWAS p values			
						Overall vs. controls	ER+ vs. controls	ER− vs. controls	ER+ vs. ER−
q13.2	ZNF404	19: 44376515–44388203	445	0.31	0.06	**2.0E–13**	**5.1E–12**	**1.7E–08**	6.9E–01
	ZNF155	19: 44472014–44502477	486	0.39	0.11	**8.8E–09**	**1.0E–07**	**2.0E–05**	3.3E–01
	RP11–15A1.7	19: 44501048–44506988	477	0.29	0.14	**9.7E–12**	**7.3E–10**	**5.1E–07**	5.5E–01
q11.23	CPNE1	20: 34213953–34220170	299	0.25	0.20	**5.3E–05**	3.3E–04	1.4E–01	4.9E–01

Note: Significant associations after Bonferroni adjustment ($p < .05/901$) are in bold.

Abbreviations: ER, estrogen receptor; SNP, single nucleotide polymorphisms; TWAS, transcriptome-wide association studies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 2

Summary of conditional analysis at known breast cancer risk region

	Before conditional analysis			After conditional analysis				
Gene	Number of SNPs	Number of significant SNPs	Index GWAS SNP p value	Number of significant SNPs	Index SNP	Smallest conditional p values	Ratio[a]	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–05
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	9.90E–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
STXBP4	687	46	2E–28	0	rs244353	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.

[a] Proportion 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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TABLE 3

Conditional and joint analysis of gene region with multiple TWAS significant genes

Region[a]	Gene (colocalized)	Marginal TWAS		COJO	
		Z score	*p* Value	*Z* score	*p* Value
2q33	**ALS2CR12**	6.7	2.15E–11	4.6	3.70E–06
	CASP8	–5.22	1.76E–07	–2	5.00E–02
5q14	**ATG10**	–6.37	1.85E–10	–6.37	1.85E–10
	ATP6AP1L	–5.18	2.25E–07	–0.85	0.4
6q22	RP11-73O6.3	–6.92	4.46E–12	0.18	0.86
	L3MBTL3	–7.46	8.45E–14	–7.46	8.45E–14
15q24	**ULK3**	–4.11	3.87E–05	–4.1	3.90E–05
	MAN2C1	–4.95	7.37E–07	–5	7.40E–07
	CTD-2323K18.1	–4.95	7.49E–07	–1.7	0.083
17q21	LRRC37A4P	6.29	3.12E–10	0.25	0.8
	CRHR1-IT1	–6.3	2.91E–10	–6.3	2.91E–10
	CRHR1	–5.46	4.84E–08	–0.28	0.78
	KANSL1-AS1	–6.28	3.37E–10	–0.04	0.97
	LRRC37A	–5.37	7.89E–08	1.83	0.07
	LRRC37A2	–5.12	3.07E–07	1.81	0.07
19q13	**ZNF404**	7.35	2.04E–13	3.5	0.001
	ZNF155	5.75	8.81E–09	–2	0.042
	RP11-15A1.7	6.81	9.67E–12	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; sincesince 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.