View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Apollo

Full title: Assessment of interactions between 205 breast cancer susceptibility loci and 13 established risk factors in relation to breast cancer risk in the Breast Cancer Association Consortium

Short title: Gene-environment interactions and breast cancer risk

Authors and Affiliations:

Pooja Middha Kapoor[1, 2], Sara Lindström[3, 4], Sabine Behrens[1], Xiaoliang Wang[3, 4], Kyriaki Michailidou[5, 6], Manjeet K. Bolla[5], Qin Wang[5], Joe Dennis[5], Alison M. Dunning[113], Paul D.P. Pharoah[5, 113], Marjanka K. Schmidt[29, 114], Peter Kraft[46, 70], Montserrat García-Closas[7, 115], Douglas F. Easton[5, 113], Roger L. Milne[43, 56, 100], Jenny Chang-Claude[1, 87] on behalf of the Breast Cancer Association Consortium*

* Remaining list of authors elsewhere

Keywords: Gene-environment interaction, Breast cancer, Single nucleotide polymorphism, Epidemiology, Risk factors, Europeans

Abstract – 250 words

Main text – 3502 words

Corresponding author:

Prof. Dr. Jenny Chang-Claude

Unit of Genetic Epidemiology, Division of Cancer Epidemiology

German Cancer Research Center (DKFZ)

Im Neuenheimer Feld 581, 69120 Heidelberg, Germany

Tel: +49-6221 42 2373, Fax: +49-6221 42 2203

E-mail: j.chang-claude@dkfz.de

Abstract

Background: Previous gene-environment interaction studies of breast cancer risk have provided sparse evidence of interactions. Using the largest available dataset to date, we performed a comprehensive assessment of potential effect modification of 205 common susceptibility variants by 13 established breast cancer risk factors including replication of previously reported interactions.

Methods: Analyses were performed using 28,176 cases and 32,209 controls genotyped with iCOGS array and 44,109 cases and 48,145 controls genotyped using OncoArray from the Breast Cancer Association Consortium (BCAC). Gene-environment interactions were assessed using unconditional logistic regression and likelihood ratio tests for breast cancer risk overall and by estrogen-receptor(ER) status. Bayesian False Discovery Probability was used to assess the noteworthiness of the meta-analyzed array-specific interactions.

Results: Noteworthy evidence of interaction at ≤1% prior probability was observed for three SNP-risk factor pairs. SNP rs4442975 was associated with a greater reduced risk of ER-positive breast cancer (OR_{int} = 0.85 (0.78 – 0.93), $p_{int} = 2.8 \times 10^{-4}$) and overall breast cancer (OR_{int} = 0.85 (0.78 – 0.92), $p_{int} = 7.4 \times 10^{-5}$) in current users of estrogen-progesterone therapy compared to non-users. This finding was supported by replication using OncoArray data of the previously reported interaction between rs13387042 (r^2 = 0.93 with rs4442975) and current estrogen-progesterone therapy for overall disease (p_{int} = 0.004). The two other interactions suggested stronger associations between SNP rs6596100 and ER-negative breast cancer with increasing parity and younger age at first birth.

Conclusion: Overall, our study does not suggest strong effect modification of common breast cancer susceptibility variants by established risk factors.

Key messages

- The association between common breast cancer susceptibility loci and breast cancer risk is not strongly modified by established breast cancer risk factors.
- The combined effect of susceptibility loci and established risk factors is thus well described by a multiplicative model.
- We found one noteworthy G x E interaction with overall and ER-positive breast cancer risk, which was replicated, and two novel noteworthy G x E interactions with ER-negative breast cancer risk.
- In an independent dataset, we replicated two previously reported G x E interactions.

Introduction

Breast cancer is a complex disease with both environmental and genetic factors contributing to risk. Well-established modifiable and non-modifiable environmental factors include age at menarche, parity, age at first birth, breastfeeding, body mass index (BMI), use of menopausal hormonal therapy (MHT), and alcohol consumption (1-6). In addition, high to moderate-risk gene mutations such as *BRCA1, BRCA2, TP53, ATM,* and *CHEK2* increase the risk of breast cancer (7-14), as well as multiple common, low-risk single nucleotide polymorphisms (SNPs) discovered through genome-wide association studies (GWAS). Approximately 170 genome-wide significant breast cancer susceptibility loci have been identified, including the recently published 65 novel loci associated with overall breast cancer and 10 loci with estrogen receptor (ER)-negative breast cancer risk, identified through the OncoArray project (15, 16).

Estimation of any combined effect of genetic and environmental factors, including gene-environment (G x E) interactions is considered to possibly improve breast cancer risk prediction, and hence identification of women at high-risk for targeted prevention. However, development of these risk models depends on knowledge of the joint effects of genetic and environmental risk factors, in particular departures from a multiplicative model (that is, G x E interaction on relative risk scale) (17). More importantly, G x E studies of individual susceptibility loci may also provide insight on potential underlying biological mechanisms that could mediate causal effects of a factor on risk of breast cancer.

Previous G x E interaction studies of breast cancer have reported nearly 30 potential G x E interactions with little evidence of departures from multiplicative model (18, 19). Most reported G x E interactions for breast cancer have not been replicated in independent datasets. Two G x E

interactions were replicated using data from the Breast Cancer Association Consortium (BCAC) (20), but were not replicated in a smaller study by the Breast and Prostate Cancer Cohort Consortium (21). In this study, we assess interactions between 205 known common breast cancer susceptibility loci and 13 established environmental risk factors in relation to risk of overall and estrogen receptor (ER)-specific breast cancer for women of European ancestry, using the largest available dataset to date from the Breast Cancer Association Consortium (BCAC). Additionally, we attempted to replicate previously reported potential G x E interactions (18).

Materials and Methods

Study population

We analyzed data from 46 studies (16 prospective cohorts, 14 population-based case-control studies and 16 non-population based studies) participating in BCAC (**Supplementary Table 1**). Participants were excluded if they were male, were of non-European descent, had breast tumors of unknown invasiveness, or had in-situ disease or prevalent disease at the time of assessment. Women with unknown age at reference date (defined as date of diagnosis for cases and interview for controls) were also excluded. For each risk factor, only studies with risk factor information for at least 150 cases and 150 controls were included. All participating studies were approved by the relevant ethics committees and informed consent was obtained from study participants.

Data harmonization and variable definition

Data for risk factors from different studies were harmonized according to a common data dictionary and centrally quality controlled. For both case-control and cohort studies, epidemiological risk factor data was derived with reference to reference date (described above). We used reference age as surrogate to categorize women as probably pre-menopausal (<54

years) or post-menopausal (≥54 years) status. The environmental variables available for analysis were: age at menarche (per 2 years), ever parous (yes or no), and for parous women, number of full-term pregnancies (1, 2, 3 and ≥4), age at first full-term pregnancy (per 5 years), ever breastfed (yes or no), duration of breastfeeding (per 12 months), and for all women, ever use of oral contraceptives (yes or no), adult body mass index (BMI) separately for pre- and postmenopausal women (per 5 kg/m^2), adult height (per 5 cm), lifetime alcohol consumption (per 10 g/day), current smoking (yes or no), and current use of combined estrogen-progesterone menopausal hormonal therapy (MHT) (yes or no) as well as current use of estrogen-only MHT for postmenopausal women (yes or no).

Genetic data

Samples were genotyped using one of the two SNP arrays – iCOGS(22) or OncoArray(15). Included in the analyses were 28 176 cases and 32 209 controls of European ancestry genotyped by the custom iSelect genotyping array (iCOGS), comprising 211 155 SNPs(22), and 44 109 cases and 48 145 controls genotyped using the OncoArray 500K, comprising 533 000 SNPs, nearly 260 000 of which were selected as a "GWAS backbone" (Illumina HumanCore) (23). These data were used to impute genotypes for ~11.8M SNPs using the 1000 Genomes Project (phase 3 version 5) reference panel (15, 16). Details of genotyping and quality control procedures for the iCOGS and OncoArray projects are described in more detail elsewhere (15, 22, 23).

A total of 205 common breast cancer susceptibility variants were selected for evaluation of G x E interactions (**Supplementary Table 2**). These variants have been associated with breast cancer risk either through GWAS (24-34) or by fine mapping of associated regions (35-52). Of these,

72 were identified through the OncoArray project and had not been previously evaluated for G x E interactions (15, 16).

For replication of the previously reported interactions, we analyzed a subset of 30 544 cases and 37 616 controls genotyped using the OncoArray array, which had not been included in previous G x E studies. We evaluated 33 potential G x E interactions that had been previously reported (**Supplementary Table 3**) (18).

Statistical analysis

Unconditional logistic regression analysis was employed to assess associations of SNPs and risk factors with breast cancer risk. For SNPs, the estimated number of minor alleles based on imputation was included as a continuous variable. SNP-risk factor interactions were assessed using likelihood ratio tests, based on unconditional logistic regression models with and without an interaction term between the SNP and risk factor of interest. All analyses were adjusted for study, reference age, and ten ancestry-informative principal components. To account for differential main effects of risk factors by study design, we included an interaction term between the risk factor of interest and an indicator variable for study design (population-based and non-population-based), along with the main effect for study design.

Analyses were conducted separately for overall breast cancer risk and for ER-subtype specific breast cancer risk. The analyses were performed separately for women genotyped by iCOGS or OncoArray and the results were meta-analyzed using a fixed-effects inverse-variance weighted model. Between-study heterogeneity in the G x E interaction effect estimates was assessed by Cochrane's Q-test and I^2index.

MHT was classified into estrogen-progesterone therapy (EPT) and estrogen-only therapy (ET). Models assessing the association with current MHT use by type were adjusted for former use of MHT and use of any MHT preparation other than the one of interest. All analyses of MHT use were restricted to postmenopausal women. Models evaluating the association with current smoking were adjusted for former smoking.

To assess the noteworthiness of the observed G x E interactions we calculated Bayesian False Discovery Probability (BFDP) at five different prior probabilities for a true association (20%, 10%, 1%, 0.1% and 0.01%). G x E interactions with BFDP <80% were considered as noteworthy. This was based on the assumption of a four-fold cost of a false non-discovery compared with the cost of a false discovery and that the probability of observing a true interaction odds ratio (OR) inside the range of 0.66-1.50 was 95%, as proposed by Wakefield *et al*.(53). We also computed a complementary measure to BFDP known as approximate Bayes factor (ABF). It approximates the ratio of the probability of the data given that the null hypothesis is true to the probability of the data when the alternative hypothesis is true, the null hypothesis being absence of any interaction. Therefore, a lower ABF favors the alternative hypothesis over the null hypothesis of absence of an interaction. For noteworthy G x E interactions, we performed stratified analyses by categories of the environmental risk factor using logistic regression. Analyses were carried out using SAS 9.4 or R version 3.4.2. Meta-analyses and tests of between-study heterogeneity were conducted using the R package "meta" (version 4.9-2).

Results

The studies included in this analysis are listed in **Supplementary Table 1**. The number of cases and controls with data for each risk factor varied, ranging from 23 755 cases and 30 153 controls with data for parity to 5078 cases and 6867 controls with data for cumulative lifetime intake of alcohol in the iCOGS dataset and from 37 863 cases and 44 533 controls with data for parity to 12 213 cases and 13 232 controls with data for lifetime alcohol intake in the OncoArray dataset **(Supplementary Table 4 & 5**).

The SNP associations with risk of overall as well as ER-subtype breast cancer were consistent with those reported in literature (15, 16) (**Supplementary Table 2 & 3**). The associations of the environmental risk factors with breast cancer risk were as expected in the population-based studies; in brief, age at menarche, being parous, number of full-term pregnancies, ever breastfeeding, cumulative duration of breastfeeding, and premenopausal BMI were negatively associated with breast cancer risk, whereas age at first full-term pregnancy, ever use of oral contraceptives, postmenopausal BMI, current use of EPT, adult height, current smoking and cumulative alcohol consumption were all positively associated with breast cancer risk (**Table 1 & Supplementary Figures 1-3**).

We identified three SNP-risk factor interactions as noteworthy (BFDP < 0.8) at ≤1 % prior probability (**Table 2**). The strongest G x E interaction was found for SNP rs4442975 and current use of EPT ($OR_{meta\text{-}int}$ = 0.85, 95% CI = 0.78 – 0.92, $p_{meta\text{-}int}$ = 7.4 x 10^{-5}, BFDP = 0.73) with overall breast cancer at 0.1% prior probability. The minor allele of SNP rs4442975 was associated with a stronger reduced risk of breast cancer for current users of EPT (OR_{meta} = 0.74, 95% CI = 0.69 – 0.80) than for never users of MHT (OR_{meta} = 0.87, 95% CI = 0.84 – 0.90) (**Figure 1A**). This interaction was also found to be noteworthy at 1% prior probability for risk of ER-positive breast cancer ($OR_{meta\text{-}int}$ = 0.85, 95% CI = 0.78 – 0.93, $p_{meta\text{-}int}$ = 2.8 x 10^{-4}, BFDP =

0.46). The association of rs4442975 with reduced risk of ER-positive breast cancer was stronger for current users of EPT (OR_{meta} = 0.73, 95% CI = 0.68 – 0.79) than for never MHT users (OR_{meta} = 0.86, 95% CI = 0.83 – 0.89) (**Figure 1B**).

The two other noteworthy SNP-risk factor interactions were found for ER-negative breast cancer risk. The interaction between rs6596100 and number of full-term pregnancies was noteworthy at 1% prior probability ($OR_{meta\text{-}int}$ = 0.91, 95% CI = 0.85 – 0.96, $p_{meta\text{-}int}$ = 8.2 x 10^{-4}, BFDP = 0.74). The minor allele of the rs6596100 variant was associated with a reduced risk of overall breast cancer (OR_{meta} = 0.96, 95% CI = 0.94 – 0.98) and ER-positive breast cancer (OR_{meta} = 0.94, 95% CI = 0.92 – 0.96), respectively, but not ER-negative breast cancer (OR_{meta} = 1.01, 95% CI = 0.97 – 1.05). The rs6596100 associated risk of ER-negative breast cancer appears to decrease with number of full-term pregnancies for parous women, with the estimated per-allele OR_{meta} being 1.06 (95% CI = 0.95 – 1.17) for women who had had one full-term pregnancy and 0.92 (95% CI = 0.82– 1.04) for women who had had four or more full-term pregnancies (**Figure 1C**).

For parous women, we observed noteworthy evidence that the ER-negative breast cancer risk associated with rs6596100 was also modified by age at first full-term pregnancy ($OR_{meta\text{-}int}$ = 1.12, 95% CI = 1.05 – 1.19, $p_{meta\text{-}int}$ = 3.3 x 10^{-4}, BFDP = 0.56). The risk conferred by rs6596100 on ER-negative breast cancer was decreased for women with age at first full-term pregnancy below 20 years (OR_{meta} of 0.90 (95% CI = 0.79 – 1.03)) but increased for women with age at first full term pregnancy ≥ 30 years (OR_{meta} of 1.10 (95% CI = 0.97 – 1.24)) (**Figure 1D**). However, we observed between-study heterogeneity for the interaction between rs6596100 and age at first full-term pregnancy (**Supplementary Figure 4**). Several other interactions were found to be noteworthy (BFDP <0.8) at 5% prior probability (**Supplementary Table 6**). Meta-analyzed

results of all the G x E interactions for overall and ER-subtype risk are shown in **Supplementary Tables 7-9**.

In replication analyses, we found evidence for two previously reported associations in the independent subset of OncoArray data (**Supplementary Table 10**). We estimated an interaction OR for overall breast cancer of 0.80 (95% CI = 0.69-0.93, p_{int} = 0.004) for current EPT use and rs13387042, a SNP for which we had previously reported an interaction OR of 0.83 (95% CI = 0.74-0.94, $p_{int} = 2.43 \times 10^{-3}$) (20). SNP rs13387042 is in strong linkage disequilibrium with rs4442975; hence this result is consistent with the interaction observed for rs4442975 in the full dataset. In addition, we also observed evidence for a G x E interaction between rs941764 and cumulative lifetime intake of alcohol (<20 g/day vs. ≥20g/day) with ER-negative breast cancer risk (OR_{int} of 0.64, 95% CI = 0.45 – 0.92, p_{int} = 0.01), compared with OR_{int} of 0.53 (95% CI = 0.36 – 0.76, $p_{int} = 6.8 \times 10^{-4}$) in Rudolph *et al.* (54). The corresponding meta-analyzed interaction OR (per 10g/day cumulative lifetime alcohol intake) based on OncoArray and iCOGS datasets was 0.90 (95% CI = 0.81 – 0.99, p_{int} = 0.03). For the G x E interaction between SNP rs3817198 and number of children for parous women, which had the strongest evidence for overall risk of breast cancer in previous analyses (OR_{int} of 1.06 (95% CI =1.04 – 1.08), $p_{int} = 2.4 \times 10^{-06}$) (20), there was weak evidence of interaction, but in the opposite direction in the replication analyses (OR_{int} of 0.94 (95% CI = 0.94 – 1.00, p_{int} = 0.03).

Discussion

In this study, we evaluated all known common susceptibility loci for interactions with breast cancer risk factors, and found little evidence for departures from a multiplicative model. We refer to G x E interactions as effect modification conferred by epidemiological risk factors on the

association between SNPs and breast cancer risk but, it can very well be SNPs modifying the association of risk factors with breast cancer risk. We identified three noteworthy (BFDP <0.8) G x E interactions related to breast cancer risk based on prior probabilities ≤1%. The strongest evidence was found for effect modification between rs4442975 and current use of EPT with overall and ER-positive breast cancer risk. Moreover, we found evidence of interactions between the SNP rs6596100 and number of full-term pregnancies and age at first full-term pregnancy, respectively, for ER-negative breast cancer risk.

The SNP rs4442975 is located in an intergenic region on the long arm of chromosome 2 (2q35). Another SNP within the same genomic region, rs13387042, was previously reported to show an interaction also with current use of EPT (20). We replicated this interaction between rs13387042 and current use of EPT using the OncoArray dataset. The two SNPs rs13387042 and rs4442975 are highly correlated ($r^2 = 0.93$) and conditional analysis yielded a significant association only for rs4442975, so that these results reflect the same interaction. Fine-mapping and functional analyses have identified rs4442975 to be the most likely causal variant in this region (43). Thus despite the small difference in the risk estimates between never and current EPT, replication of this G x E interaction reinforced what we found previously, implicating the role of the *IGFBP5* gene and estrogen pathway in breast cancer.

Functional analyses indicate that SNP rs4442975 lies near a transcriptional enhancer which physically interacts with the *IGFBP5* promoter, suggesting that the T allele of rs4442975 decreases susceptibility to breast cancer via increased expression of insulin-like growth factor binding protein 5 (IGFBP5) (43). IGFBP5 is a key member of the insulin-like growth factor (IGF) axis which plays an important role in cellular differentiation, proliferation and apoptosis in breast cancer (55). Activation of the IGF receptors by IGF causes phosphorylation of insulin

receptor substrates (IRS-1 & IRS-2). This phosphorylation cascades multiple downstream signaling pathways such as Ras/mitogen-activated protein kinase (MAPK) and phosphoinositide (PI3K) serine-threonine kinase (AkT) which play a role in breast carcinogenesis (56, 57). Estrogen can stimulate the IGF pathway via increased expression of both insulin-like growth factor receptor-1 and IRS-1. Some studies have also reported a positive correlation between overexpression of IGFBP5 and the presence of ER in breast cancer cell lines. Progesterone has been shown to act by increasing levels of IRS-2 and sensitizing breast cancer cells to downstream signaling pathways such as MAPK and Akt (58-60). It is plausible that exogenous hormone exposure due to estrogen and progesterone therapy may affect the regulation of the IGF pathway and thereby modulate germline *IGFPB5* variant-related susceptibility to breast cancer. Note however that two other independent breast cancer risk variants in this region (tagged by rs16857609 (13) and a 1.3kb insertion/deletion (49)) are also believed to target *IGFBP5* but we did not find evidence for interactions between these variants and current EPT use.

Women of young age at first pregnancy are known to have increased circulating sex hormone binding globulin and prolactin but decreased total estrogen levels (61, 62). Likewise, women who have had multiple full-term pregnancies have an overall decreased lifetime exposure to estrogen (61, 63, 64). The association of rs6596100 with ER-negative breast cancer risk was found to be modified by number of full-term pregnancies and age at first full-term pregnancy for parous women. Based on INQUISIT (15), the target genes of rs6596100 and highly correlated SNPs are predicted to be heat shock protein family A member 4 (*HSPA4*) and AF4/FMR2 family member 4 (*AFF4)*. INQUISIT predicts *HSPA4* as the most likely target due to overlap of multiple correlated SNPs lying in *HSPA4* promoter region, distal regulatory elements and coding sequence. *HSPA4* gene is responsible for production of heat shock proteins (Hsps), particularly

those belonging to the family HSP70. The underlying mechanisms regarding the relationship between rs6596100 and these pregnancy-related risk factors are unknown at present. It is plausible that a lower estrogenic milieu due to reproductive factors may affect the formation of multi-complexes between steroid receptors like ER and heat shock proteins (HSPs), and therefore affecting signaling pathways such as Wnt, ErbB, serine/threonine and tyrosine protein kinase, which are known to be involved in breast carcinogenesis. While there is some biological plausibility regarding the observed interactions with rs6596100, the findings nevertheless could be by chance and thus require independent replication.

The SNP rs941764 is located on chromosome 14 in intron of *CCDC88C* gene (15, 22). The effect modification of rs941764 associated ER-negative breast cancer risk by lifetime intake of alcohol was first reported by Rudolph et al.(54). We replicated this G x E interaction in an independent dataset in our study. Mutations in this gene region have been associated with dysregulation of Wnt signaling in neural disorders such as congenital hydrocephalus (65). This gene codes a Hook-related protein (HkRP2) that binds to an important scaffold protein, Dishevelled, in the Wnt signaling pathway, affecting all downstream activity (65).

A role of alcohol has been well recognized in initiation and progression of breast cancer presumably via multiple cellular and molecular mechanisms, including the EGFR/ErbB2 pathways. Downstream to EGFR/ErbB2 pathways lie multiple pathways such as the MAPK, Wnt/GSK3β/β-catenin pathways (66). Therefore, alcohol consumption could affect the risk of ER-negative breast cancer through dysregulation of Wnt signaling.

Our study provides the most comprehensive evaluation to date of potential effect modification of all known common genetic susceptibility variants by environmental risk factors for breast

cancer. Our findings are based on the largest available dataset on breast cancer. Despite its large sample size, the study may remain statistically underpowered, considering the rather modest effect sizes of most of the common variants associated with breast cancer risk, and particularly for risk factors for which we have less data (Supplementary table 11) (18). Statistical power was further diminished for subtype-specific analyses due to reduced sample sizes, especially for ER-negative breast cancer (10,896 ER-negative cases in the combined iCOGS and OncoArray dataset) (18). The lack of strong effect modifications for breast cancer could also be explained by the overall weak to moderate associations of environmental risk factors except for MHT use with breast cancer risk along with the modest associations of common genetic variants. A further limitation of our study is that the findings may not be generalizable to other racial/ethnic groups since the analyses were restricted to women of European ancestry.

In conclusion, our analyses suggest that most of the associated effects of breast cancer susceptibility loci and environmental risk factors are consistent with a multiplicative model. The strongest evidence for an interaction was between the candidate causal variant rs4442975 at 2q35 and current use of EPT. The associated effect is supported by a plausible underlying biological mechanism, but further epidemiological and functional validation will be required to determine whether the interaction is genuine. The newly reported results for ER-negative breast cancer risk generate plausible biological hypotheses and may inform future functional studies. Overall, the results from our analyses do not suggest strong effect modification of the association between breast cancer susceptibility loci and risk of breast cancer by established epidemiological risk factors.

Remaining list of authors from the Breast Cancer Association Consortium

Thomas Ahearn[7], Irene L. Andrulis[8, 9], Hoda Anton-Culver[10], Volker Arndt[11], Kristan J. Aronson[12], Paul L. Auer[13, 14], Annelie Augustinsson[15], Laura E. Beane Freeman[7], Matthias W. Beckmann[16], Javier Benitez[17, 18], Leslie Bernstein[19], Takiy Berrandou[20], Stig E. Bojesen[21-23], Hiltrud Brauch[24-26], Hermann Brenner[11, 26, 27], Ian W. Brock[28], Annegien Broeks[29], Angela Brooks-Wilson[30, 31], Katja Butterbach[11], Qiuyin Cai[32], Daniele Campa[1, 33], Federico Canzian[34], Brian D. Carter[35], Jose E. Castelao[36], Stephen J. Chanock[7], Georgia Chenevix-Trench[37], Ting-Yuan David Cheng[38], Christine L. Clarke[39], Emilie Cordina-Duverger[20], Fergus J. Couch[40], Angela Cox[28], Simon S. Cross[41], Kamila Czene[42], James Y. Dai[13], Gillian S. Dite[43], H. Shelton Earp[44], A. Heather Eliassen [45, 46], Mikael Eriksson[42], D. Gareth Evans[47, 48], Peter A. Fasching[16, 49], Jonine Figueroa[7, 50, 51], Henrik Flyger[52], Lin Fritschi[53], Marike Gabrielson[42], Manuela Gago-Dominguez[54, 55], Susan M. Gapstur[35], Mia M. Gaudet[35], Graham G. Giles[43, 56, 57], Anna González-Neira[18], Anne Grundy[58], Pascal Guénel[20], Lothar Haeberle[59], Christopher A. Haiman[60], Niclas Håkansson[61], Per Hall[42, 62], Ute Hamann[63], Susan E. Hankinson[45, 64], Elaine F. Harkness[65-67], Tricia Harstad[68], Wei He[42], Jane Heyworth[69], Robert N. Hoover[7], John L. Hopper[43], Keith Humphreys[42], David J. Hunter[46, 70, 71], ABCTB Investigators[72], kConFab/AOCS Investigators[73, 74], Pablo Isidro Marrón[75], Esther M. John[76], Michael E. Jones[77], Audrey Jung[1], Rudolf Kaaks[1], Renske Keeman[29], Cari M. Kitahara[78], Yon-Dschun Ko[79], Stella Koutros[7], Ute Krüger[15], Diether Lambrechts[80, 81], Loic Le Marchand[82], Eunjung Lee[60], Flavio Lejbkowicz[83], Martha Linet[78], Jolanta Lissowska[84], Ana Llaneza[85], Wing-Yee Lo[24, 25], Enes Makalic[43], Maria Elena Martinez[55, 86], Tabea Maurer[87], Victor M. Muñoz-Garzon[88], Susan L. Neuhausen[19], Patrick Neven[89], William G. Newman[47, 48], Sune F. Nielsen[21, 22], Børge G. Nordestgaard[21-23], Aaron Norman[68], Katie M. O'Brien[90], Andrew F. Olshan[91], Janet E. Olson[68], Håkan Olsson[15], Nick Orr[92], Charles M.

Perou[93], Mila Pinchev[83], Ross Prentice[13], Gad Rennert[83], Hedy S. Rennert[83], Kathryn J. Ruddy[94], Dale P. Sandler[90], Michael O. Schneider[59], Minouk J. Schoemaker[77], Ben Schöttker[11, 95], Christopher Scott[68], Rodney J. Scott[96-98], Mark E. Sherman[99], Martha J. Shrubsole[32], Xiao-Ou Shu[32], Melissa C. Southey[100, 101], John J. Spinelli[102, 103], Jennifer Stone[43, 104], Anthony J. Swerdlow[77, 105], Rulla M. Tamimi[45, 46, 70], Jack A. Taylor[90, 106], Kathrin Thöne[87], Melissa A. Troester[91], Thérèse Truong[20], Celine M. Vachon[107], Chantal van Ongeval[89], Elke M. van Veen[47, 48], Philippe Wagner[15], Clarice R. Weinberg[108], Hans Wildiers[89], Walter Willett [46, 109, 110], Stacey J. Winham[111], Alicja Wolk[61, 112], Xiaohong R. Yang[7], Wei Zheng[32], Argyrios Ziogas[10]

[1] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[2] Faculty of Medicine, University of Heidelberg, Heidelberg, Germany.

[3] Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA.

[4] Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

[5] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[6] Department of Electron Microscopy/Molecular Pathology and The Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus.

[7] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA.

[8] Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada.

[9] Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.

[10] Department of Epidemiology, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA.

[11] Division of Clinical Epidemiology and Aging Research, C070, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[12] Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada.

[13] Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

[14] Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA.

[15] Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden.

[16] Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.

[17] Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia, Spain.

[18] Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain.

[19] Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA.

[20] Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France.

[21] Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.

[22] Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.

[23] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[24] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany.

[25] University of Tübingen, Tübingen, Germany.

[26] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.

[27] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.

[28] Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK.

[29] Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

[30] Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada.

[31] Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada.

[32] Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA.

[33] Department of Biology, University of Pisa, Pisa, Italy.

[34] Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[35] Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA.

[36] Oncology and Genetics Unit, Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain.

[37] Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

[38] Division of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA.

[39] Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia.

[40] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.

[41] Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK.

[42] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[43] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.

[44] Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[45] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

[46] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[47] Division of Evolution and Genomic Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.

[48] Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester NIHR Biomedical Research Centre, Manchester University Hospitals NHS, Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

[49] David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA.

[50] Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh Medical School, Edinburgh, UK.

[51] Cancer Research UK Edinburgh Centre, Edinburgh, UK.

[52] Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.

[53] School of Public Health, Curtin University, Perth, Western Australia, Australia.

[54] Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain.

[55] Moores Cancer Center, University of California San Diego, La Jolla, CA, USA.

[56] Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia.

[57] Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia.

[58] Centre de Recherche du Centre Hospitalier de Université de Montréal (CHUM), Université de Montréal, Montréal, QC, Canada.

[59] Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany.

[60] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

[61] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

[62] Department of Oncology, Södersjukhuset, Stockholm, Sweden.

[63] Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[64] Department of Biostatistics & Epidemiology, University of Massachusetts, Amherst, Amherst, MA, USA.

[65] Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.

[66] Nightingale Breast Screening Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK.

[67] NIHR Manchester Biomedical Research Unit, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

[68] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.

[69] School of Population and Global Health, The University of Western Australia, Perth, Western Australia, Australia.

[70] Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[71] Nuffield Department of Population Health, University of Oxford, Oxford, UK.

[72] Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia.

[73] Research Department, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia.

[74] Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Victoria, Australia.

[75] Principado de Asturias Biobank, Instituto de Investigación Sanitaria del Principado de Asturias, Hospital Universitario Central de Asturias, Oviedo, Spain.

[76] Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA.

[77] Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK.

[78] Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.

[79] Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany.

[80] VIB Center for Cancer Biology, VIB, Leuven, Belgium.

[81] Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium.

[82] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA.

[83] Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel.

[84] Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Cancer Center, Oncology Institute, Warsaw, Poland.

[85] General and Gastroenterology Surgery Service, Hospital Universitario Central de Asturias, Oviedo, Spain.

[86] Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA.

[87] Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

[88] Radiation Oncology, Hospital Meixoeiro-XXI de Vigo, Vigo, Spain.

[89] Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium.

[90] Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.

[91] Department of Epidemiology, Gilliungs School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[92] Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Ireland, UK.

[93] Department of Genetics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[94] Department of Oncology, Mayo Clinic, Rochester, MN, USA.

[95] Network Aging Research, University of Heidelberg, Heidelberg, Germany.

[96] Division of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, New South Wales, Australia.

[97] Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Callaghan, New South Wales, Australia.

[98] Hunter Medical Research Institute, John Hunter Hospital, Newcastle, New South Wales, Australia.

[99] Department of Health Sciences Research, Mayo Clinic College of Medicine, Jacksonville, FL, USA.

[100] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia.

[101] Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia.

[102] Population Oncology, BC Cancer, Vancouver, BC, Canada.

[103] School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada.

[104] The Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and University of Western Australia, Perth, Western Australia, Australia.

[105] Division of Breast Cancer Research, The Institute of Cancer Research, London, UK.

[106] Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.

[107] Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA.

[108] Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.

[109] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[110] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
[111] Department of Health Science Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA.
[112] Department of Surgical Sciences, Uppsala University, Uppsala, Sweden.
[113] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.
[114] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands.
[115] Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK.

References

1. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet. 1997;350(9084):1047-59.

2. Baer HJ, Rich-Edwards JW, Colditz GA, Hunter DJ, Willett WC, Michels KB. Adult height, age at attained height, and incidence of breast cancer in premenopausal women. Int J Cancer. 2006;119(9):2231-5.

3. Hunter DJ, Colditz GA, Hankinson SE, Malspeis S, Spiegelman D, Chen W, et al. Oral contraceptive use and breast cancer: a prospective study of young women. Cancer Epidemiol Biomarkers Prev. 2010;19(10):2496-502.

4. Collaborative Group on Hormonal Factors in Breast C. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012;13(11):1141-51.

5. Jung S, Wang M, Anderson K, Baglietto L, Bergkvist L, Bernstein L, et al. Alcohol consumption and breast cancer risk by estrogen receptor status: in a pooled analysis of 20 studies. Int J Epidemiol. 2016;45(3):916-28.

6. World Cancer Research Fund International/American Institute for Cancer Research. Continuous Update Project Report: Diet, Nutrition, Physical Activity and Breast Cancer. 2017.

7. Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Am J Hum Genet. 1995;56(1):265-71.

8. Breast Cancer Linkage C. Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst. 1999;91(15):1310-6.

9. King MC, Marks JH, Mandell JB, New York Breast Cancer Study G. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 2003;302(5645):643-6.
10. Garber JE, Goldstein AM, Kantor AF, Dreyfus MG, Fraumeni JF, Jr., Li FP. Follow-up study of twenty-four families with Li-Fraumeni syndrome. Cancer Res. 1991;51(22):6094-7.
11. Birch JM, Alston RD, McNally RJ, Evans DG, Kelsey AM, Harris M, et al. Relative frequency and morphology of cancers in carriers of germline TP53 mutations. Oncogene. 2001;20(34):4621-8.
12. Rapakko K, Allinen M, Syrjakoski K, Vahteristo P, Huusko P, Vahakangas K, et al. Germline TP53 alterations in Finnish breast cancer families are rare and occur at conserved mutation-prone sites. Br J Cancer. 2001;84(1):116-9.
13. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55-9.
14. Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, et al. Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst. 2005;97(11):813-22.
15. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551(7678):92-4.
16. Milne RL, Kuchenbaecker KB, Michailidou K, Beesley J, Kar S, Lindstrom S, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet. 2017;49(12):1767-78.
17. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005;6(4):287-98.

18. Rudolph A, Chang-Claude J, Schmidt MK. Gene-environment interaction and risk of breast cancer. Br J Cancer. 2016;114(2):125-33.

19. Barrdahl M, Rudolph A, Hopper JL, Southey MC, Broeks A, Fasching PA, et al. Gene-environment interactions involving functional variants: Results from the Breast Cancer Association Consortium. Int J Cancer. 2017;141(9):1830-40.

20. Nickels S, Truong T, Hein R, Stevens K, Buck K, Behrens S, et al. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. PLoS Genet. 2013;9(3):e1003284.

21. Barrdahl M, Canzian F, Joshi AD, Travis RC, Chang-Claude J, Auer PL, et al. Post-GWAS gene-environment interplay in breast cancer: results from the Breast and Prostate Cancer Cohort Consortium and a meta-analysis on 79,000 women. Hum Mol Genet. 2014;23(19):5260-70.

22. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45(4):353-61, 61e1-2.

23. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev. 2017;26(1):126-35.

24. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007;447(7148):1087-93.

25. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009;41(5):579-84.

26. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010;42(6):504-7.

27. Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet. 2012;44(3):312-8.

28. Long J, Cai Q, Sung H, Shi J, Zhang B, Choi JY, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. PLoS Genet. 2012;8(2):e1002532.

29. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Hum Mol Genet. 2012;21(24):5373-84.

30. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013;45(4):392-8, 8e1-2.

31. Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. Nat Genet. 2014;46(8):886-90.

32. Milne RL, Burwinkel B, Michailidou K, Arias-Perez JI, Zamora MP, Menendez-Rodriguez P, et al. Common non-synonymous SNPs associated with breast cancer susceptibility:

findings from the Breast Cancer Association Consortium. Hum Mol Genet. 2014;23(22):6096-111.

33. Sawyer E, Roylance R, Petridis C, Brook MN, Nowinski S, Papouli E, et al. Genetic predisposition to in situ and invasive lobular carcinoma of the breast. PLoS Genet. 2014;10(4):e1004285.

34. Couch FJ, Kuchenbaecker KB, Michailidou K, Mendoza-Fandino GA, Nord S, Lilyquist J, et al. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. Nat Commun. 2016;7:11375.

35. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet. 2009;41(5):585-90.

36. Udler MS, Ahmed S, Healey CS, Meyer K, Struewing J, Maranian M, et al. Fine scale mapping of the breast cancer 16q12 locus. Hum Mol Genet. 2010;19(12):2507-15.

37. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet. 2011;43(12):1210-4.

38. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet. 2013;45(4):371-84, 84e1-2.

39. French JD, Ghoussaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, et al. Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am J Hum Genet. 2013;92(4):489-503.

40. Gaudet MM, Kuchenbaecker KB, Vijai J, Klein RJ, Kirchhoff T, McGuffog L, et al. Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. PLoS Genet. 2013;9(3):e1003173.

41. Long J, Delahanty RJ, Li G, Gao YT, Lu W, Cai Q, et al. A common deletion in the APOBEC3 genes and breast cancer risk. J Natl Cancer Inst. 2013;105(8):573-9.

42. Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, et al. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. Am J Hum Genet. 2013;93(6):1046-60.

43. Ghoussaini M, Edwards SL, Michailidou K, Nord S, Cowper-Sal Lari R, Desai K, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nat Commun. 2014;4:4999.

44. Darabi H, McCue K, Beesley J, Michailidou K, Nord S, Kar S, et al. Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate NRBF2 Expression. Am J Hum Genet. 2015;97(1):22-34.

45. Glubb DM, Maranian MJ, Michailidou K, Pooley KA, Meyer KB, Kar S, et al. Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. Am J Hum Genet. 2015;96(1):5-20.

46. Lin WY, Camp NJ, Ghoussaini M, Beesley J, Michailidou K, Hopper JL, et al. Identification and characterization of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. Hum Mol Genet. 2015;24(1):285-98.

47. Orr N, Dudbridge F, Dryden N, Maguire S, Novo D, Perrakis E, et al. Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. Hum Mol Genet. 2015;24(10):2966-84.

48. Darabi H, Beesley J, Droit A, Kar S, Nord S, Moradi Marjaneh M, et al. Fine scale mapping of the 17q22 breast cancer locus using dense SNPs, genotyped within the Collaborative Oncological Gene-Environment Study (COGs). Sci Rep. 2016;6:32512.

49. Dunning AM, Michailidou K, Kuchenbaecker KB, Thompson D, French JD, Beesley J, et al. Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nat Genet. 2016;48(4):374-86.

50. Ghoussaini M, French JD, Michailidou K, Nord S, Beesley J, Canisus S, et al. Evidence that the 5p12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGF10 and MRPS30 Regulation. Am J Hum Genet. 2016;99(4):903-11.

51. Lawrenson K, Kar S, McCue K, Kuchenbaeker K, Michailidou K, Tyrer J, et al. Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. Nat Commun. 2016;7:12675.

52. Wyszynski A, Hong CC, Lam K, Michailidou K, Lytle C, Yao S, et al. An intergenic risk locus containing an enhancer deletion in 2q35 modulates breast cancer risk by deregulating IGFBP5 expression. Hum Mol Genet. 2016;25(17):3863-76.

53. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet. 2007;81(2):208-27.

54. Rudolph A, Milne RL, Truong T, Knight JA, Seibold P, Flesch-Janys D, et al. Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors. Int J Cancer. 2015;136(6):E685-96.

55. Beattie J, Allan GJ, Lochrie JD, Flint DJ. Insulin-like growth factor-binding protein-5 (IGFBP-5): a critical member of the IGF axis. Biochem J. 2006;395(1):1-19.

56. Akkiprik M, Feng Y, Wang H, Chen K, Hu L, Sahin A, et al. Multifunctional roles of insulin-like growth factor binding protein 5 in breast cancer. Breast Cancer Res. 2008;10(4):212.

57. Zhang X, Yee D. Tyrosine kinase signalling in breast cancer: insulin-like growth factors and their receptors in breast cancer. Breast Cancer Res. 2000;2(3):170-5.

58. Cui X, Lazard Z, Zhang P, Hopp TA, Lee AV. Progesterone crosstalks with insulin-like growth factor signaling in breast cancer cells via induction of insulin receptor substrate-2. Oncogene. 2003;22(44):6937-41.

59. Fagan DH, Yee D. Crosstalk between IGF1R and estrogen receptor signaling in breast cancer. J Mammary Gland Biol Neoplasia. 2008;13(4):423-9.

60. Yee D, Lee AV. Crosstalk between the insulin-like growth factors and estrogens in breast cancer. J Mammary Gland Biol Neoplasia. 2000;5(1):107-15.

61. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. Epidemiol Rev. 1993;15(1):36-47.

62. Bernstein L. Epidemiology of endocrine-related risk factors for breast cancer. J Mammary Gland Biol Neoplasia. 2002;7(1):3-15.

63. Dall GV, Britt KL. Estrogen Effects on the Mammary Gland in Early and Late Life and Breast Cancer Risk. Front Oncol. 2017;7:110.

64. Fortner RT, Hankinson SE. Reproductive and Hormonal factors and Breast Cancer. Translational Endocrinology & Metabolism: Breast Cancer Update. 32012. p. 95-116.

65. Ekici AB, Hilfinger D, Jatzwauk M, Thiel CT, Wenzel D, Lorenz I, et al. Disturbed Wnt Signalling due to a Mutation in CCDC88C Causes an Autosomal Recessive Non-Syndromic Hydrocephalus with Medial Diverticulum. Mol Syndromol. 2010;1(3):99-112.

66. Wang Y, Xu M, Ke ZJ, Luo J. Cellular and molecular mechanisms underlying alcohol-induced aggressiveness of breast cancer. Pharmacol Res. 2017;115:299-308.

Supplementary Information

Supplementary Table 1 (S1): Studies participating in G x E analysis with number of cases and controls.

Supplementary Table 2 (S2): Associations between 205 common breast cancer susceptibility loci with breast cancer risk in European population, overall and by ER status.

Supplementary Table 3 (S3): Associations between 33 replication SNPs with breast cancer risk in European population, overall and by ER status.

Supplementary Table 4 (S4): Number of cases and controls for each environmental risk factor by study design in iCOGS and OncoArray dataset.

Supplementary Table 5 (S5): Number of cases and controls for each environmental risk factor by overall and ER-status in complete and replication dataset.

Supplementary Table 6 (S6): G x E interactions with BFDP <80% at 5% prior probability (meta-analyzed results).

Supplementary Table 7 (S7): Meta-analyzed G x E interactions between 205 common genetic susceptibility loci and environmental risk factors for overall breast cancer risk.

Supplementary Table 8 (S8): Meta-analyzed G x E interactions between 205 common genetic susceptibility loci and environmental risk factors for ER-positive breast cancer risk.

Supplementary Table 9 (S9): Meta-analyzed G x E interactions between 205 common genetic susceptibility loci and environmental risk factors for ER-negative breast cancer risk.

Supplementary Table 10 (S10): Interaction odds ratio (OR) and 95% confidence intervals (CI) for previously reported G x E interactions in an independent dataset.

Supplementary Table 11 (S11): Power for detecting different gene-environment interaction effect estimates (OR of 0.75 to 1.50) given different minor allele frequencies (0.05 to 0.45) for 1:1 unmatched case-control study. Power calculation is performed by Quanto 1.2.4, assuming a population prevalence of disease of 1%, 15% prevalence of the environmental factor. We assumed a log-additive inheritance model with marginal effect estimate of SNP = 1.10 and marginal effect estimate of environmental factor = 1.20 and two-sided alpha of 5×10^{-8}.

Supplementary Figure 1: Forest plot of meta-analyzed study-wise odds ratios and 95% confidence intervals of population-based studies for associations between environmental risk factors and overall breast cancer risk

Supplementary Figure 2: Forest plot of meta-analyzed study-wise odds ratios and 95% confidence intervals of population-based studies for associations between environmental risk factors and ER-positive breast cancer risk

Supplementary Figure 3: Forest plot of meta-analyzed study-wise odds ratios and 95% confidence intervals of population-based studies for associations between environmental risk factors and ER-negative breast cancer risk

Supplementary Figure 4: Forest plot of meta-analyses of study-wise odds ratios and 95% confidence intervals for G x E interactions between SNPs and environmental risk factors of breast cancer (from Table 2) separately for OncoArray and iCOGS datasets.

Funding and Acknowledgments

Funding

This work was supported by following funding agencies.

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number 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 n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, 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 a 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 AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute

(grant number Z01-CP010119), and the National Institute of Environmental Health Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. 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). The BREast Oncology GAlician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, 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. CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. 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). The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The CTS was supported by the California Breast Cancer Act of 1993, the California Breast Cancer Research Fund (contract 97-10500) and the National Institutes of Health (R01 CA77398, K05 CA136967, UM1 CA164917, and U01 CA199277). 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 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). 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 GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. 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, 199600). G.C.T. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. 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]. The MCBCS was supported by the

NIH grants CA192393, CA116167, CA176785 an 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. MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057 and 396414, and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. The MEC was support 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. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. 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,

U01 CA179715), and the North Carolina University Cancer Research Fund. 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 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. PROCAS is funded from NIHR grant PGfAR 0707-10031. Prof. Dr. D Gareth Evans is supported by the NIHR Biomedical Research Centre in Manchester (IS-BRC-1215-20007). 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 SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank. SEARCH is funded by Cancer Research UK [C490/A10124, 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. 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 SMC is funded by the Swedish Cancer Foundation. The UCIBCS component of this research was supported by the NIH [CA58860, CA92044] and the Lon V Smith Foundation [LVS39420]. The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges

NHS funding to the NIHR Biomedical Research Centre. The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. 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, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01.

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. The COGS study would not have been possible without the contributions of the following: Per Hall (COGS); Douglas F. Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Douglas F. Easton, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. ABCS thanks the Blood bank Sanquin, The

Netherlands. ABCTB Investigators: Christine Clarke,Rosemary Balleine, Robert Baxter,Stephen Braye, Jane Carpenter, Jane Dahlstrom, John Forbes, Soon Lee, Debbie Marsh, Adrienne Morey, Nirmala Pathmanathan, Rodney Scott, Allan Spigelman, Nicholas Wilcken, Desmond Yip. Samples are made available to researchers on a non-exclusive basis. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, Rachel Lloyd. The BCINIS study would not have been possible without the contributions of Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. The BREOGAN study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREOGAN), José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS; Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain. CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. CGPS thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the

collection of blood samples for the cases. CNIO-BCS thanks Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO) and Principado de Asturias BioBank (PT17/0015/0023). Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. CTS Investigators include Leslie Bernstein, Susan Neuhausen, James Lacey, Sophia Wang, and Huiyan Ma at the Beckman Research Institute of City of Hope, Jessica Clague DeHart at the School of Community and Global Health Claremont Graduate University, Dennis Deapen, Rich Pinder, and Eunjung Lee at the University of Southern California, Pam Horn-Ross, Christina Clarke Dur and David Nelson at the Cancer Prevention Institute of California, Peggy Reynolds, at the Department of Epidemiology and Biostatistics, University of California San Francisco, Hoda Anton-Culver, Argyrios Ziogas, and Hannah Park at the University of California Irvine, and Fred Schumacher at Case Western University. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach. The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [HB, WYL], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) [HB], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [YDK, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches

Krebsforschungszentrum (DKFZ), Heidelberg, Germany [UH], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth].KARMA and SASBAC thank the Swedish Medical Research Counsel. kConFab/AOCS wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. MARIE thanks Petra Seibold, Dieter Flesch-Janys, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. MBCSG (Milan Breast Cancer Study Group): Paolo Radice, Paolo Peterlongo, Siranoush Manoukian, Bernard Peissel, Roberto Villa, Cristina Zanzottera, Bernardo Bonanni, Irene Feroce, and the personnel of the Cogentech Cancer Genetic Test Laboratory. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. NBHS thank study participants and research staff for their contributions and commitment to the studies. We would like to thank the participants and staff of the Nurses' Health Study and Nurses' Health Study II for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for

analyses and interpretation of these data. OFBCR thanks Teresa Selander, Nayana Weerasooriya. PBCS thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. PROCAS thanks NIHR for funding. SBCS thanks Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed. We thank the SEARCH and EPIC teams. UCIBCS thanks Irene Masunaka. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible.

Conflict of Interest: none declared

Table 1: Main effects for the epidemiologic variables included in the analyses, derived from population-based studies only.

Environmental risk factor	Overall breast cancer risk		ER-positive breast cancer risk		ER-negative breast cancer risk	
	Cases/Controls	OR (95% CI)	Cases/Controls	OR (95% CI)	Cases/Controls	OR (95% CI)
Age at menarche (per 2 years)	36893/46854	0.91 (0.89-0.92)	26630/46854	0.91 (0.89-0.93)	4255/25233	0.89 (0.85-0.93)
Ever parous (yes/no)	37242/47173	0.81 (0.77-0.84)	26937/47173	0.78 (0.74-0.81)	4309/25585	0.94 (0.85-1.04)
Number of full-term pregnancies (1,2,3, ≥4)	31390/41215	0.87 (0.85-0.88)	22720/41215	0.86 (0.84-0.87)	3273/18267	0.90 (0.86-0.94)
Age at first full-term pregnancy (per 5 years)[1]	30168/39850	1.14 (1.12-1.16)	21869/39850	1.17 (1.14-1.19)	3472/21422	1.02 (0.97-1.06)
Ever breastfed (yes/no)[1]	27786/30582	0.91 (0.88-0.95)	19691/30582	0.92 (0.88-0.96)	3533/19606	0.96 (0.88-1.03)
Duration of breastfeeding (per 12 months)[1]	24553/25524	0.96 (0.93-0.98)	17355/25524	0.95 (0.93-0.98)	3315/18012	0.98 (0.94-1.03)
Adult height (per 5 cm)	35767/46506	1.09 (1.08-1.10)	25763/46506	1.10 (1.09-1.12)	3954/24342	1.03 (1.00-1.05)
Premenopausal BMI (per 5 kg/m^2)	7994/10066	0.95 (0.92-0.98)	4835/9490	0.92 (0.89-0.95)	913/2030	1.07 (0.98-1.16)
Postmenopausal BMI (per 5 kg/m^2)	27495/32495	1.07 (1.05-1.09)	20503/32283	1.07 (1.05-1.09)	1758/11859	1.05 (1.00-1.11)
Ever use of oral contraceptives (yes/no)	35126/44608	1.22 (1.18-1.26)	25271/44608	1.24 (1.20-1.29)	3939/24225	1.14 (1.05-1.23)
Current use of EPT (yes/no)[2,3]	16637/17946	1.75 (1.65-1.87)	12566/17946	1.93 (1.81-2.06)	1190/7353	1.11 (0.92-1.34)
Current use of ET (yes/no)[2,3]	16444/17920	1.10 (1.03-1.17)	11829/16844	1.11 (1.03-1.19)	936/6262	1.35 (1.11-1.64)
Lifetime intake of alcohol (per 10 g/day)	15827/18723	1.07 (1.05-1.10)	11302/18723	1.09 (1.07-1.11)	1612/11562	1.03 (0.98-1.08)
Current smoking (yes/no)[4]	33737/43222	1.18 (1.13-1.24)	24123/43222	1.18 (1.12-1.25)	3707/22573	1.06 (0.96-1.18)
Pack years smoked (per 10 pack-years)[5]	7975/11709	1.02 (1.00-1.04)	5944/11709	1.02 (1.00-1.04)	896/6400	1.00 (0.95-1.04)

ER: Estrogen receptor, OR: odd ratio, CI: confidence interval, BMI: Body mass index, EPT: Estrogen-Progesterone menopausal hormonal therapy, ET: Estrogen-only menopausal hormonal therapy

All models were adjusted for reference age and study

[1] for parous women

[2] for postmenopausal women

[3] Additionally, models were adjusted for former use of menopausal hormonal therapy and use of any other menopausal hormonal therapy preparations

[4] Additionally, model was adjusted for former smoking

[5] for ever smokers

Table 2: Gene-environment interactions with Bayesian False Discovery Probability (BFDP) <80% at ≤1% prior probability.

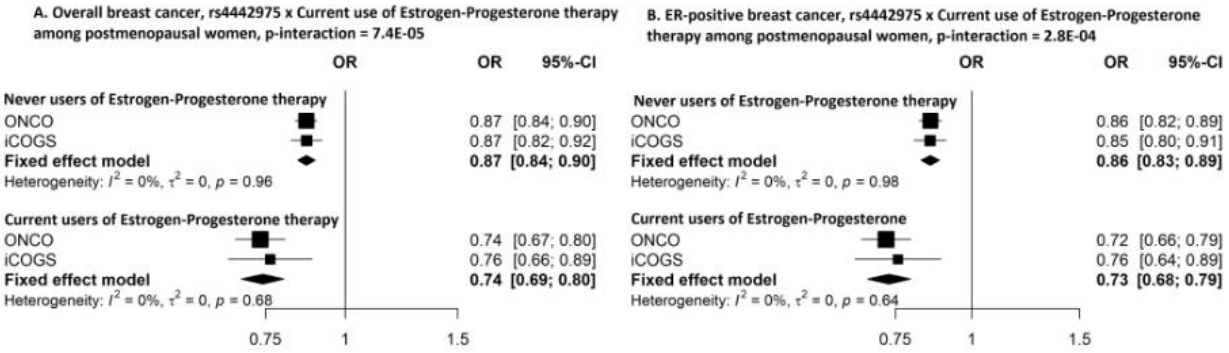
Environmental risk factor	SNP (Gene)	iCOGS OR_{int} (95% CI)	OncoArray OR_{int} (95% CI)	Meta-analysis OR_{int} (95% CI)	Meta-analysis p_{int}	Prior probability (BFDP) 0.2	0.1	0.01	0.001	0.0001	ABF
OVERALL BREAST CANCER RISK											
Current EPT use[1]	**rs4442975 (*IGFBP5*)**	0.88 (0.75 – 1.03)	0.83 (0.76 – 0.92)	0.85 (0.78 – 0.92)	7.4E-05	0.011	0.023	0.209	0.727	0.964	0.003
ER-POSITIVE BREAST CANCER RISK											
Current EPT use[1]	**rs4442975 (*IGFBP5*)**	0.89 (0.75 – 1.06)	0.84 (0.75 – 0.93)	0.85 (0.78 – 0.93)	2.8E-04	0.033	0.072	0.462	0.896	0.989	0.009
ER-NEGATIVE BREAST CANCER RISK											
Number of full-term pregnancies[2,3]	**rs6596100 (*HSPA4*)**	0.84 (0.75 – 0.93)	0.94 (0.87 – 1.01)	0.91 (0.85 – 0.96)	8.2E-04	0.104	0.207	0.742	0.967	0.997	0.029
Age at FFTP[2]	**rs6596100 (*HSPA4*)**	1.13 (1.02 – 1.26)	1.11 (1.03 – 1.19)	1.12 (1.05 – 1.19)	3.3E-04	0.048	0.103	0.558	0.927	0.992	0.012

ER: Estrogen receptor, OR_{int}: Interaction odds ratio, CI: Confidence interval, SNP: Single nucleotide polymorphism, ABF: Approximate Bayes Factor, EPT: Estrogen-Progesterone therapy, FFTP: First full-term pregnancy,

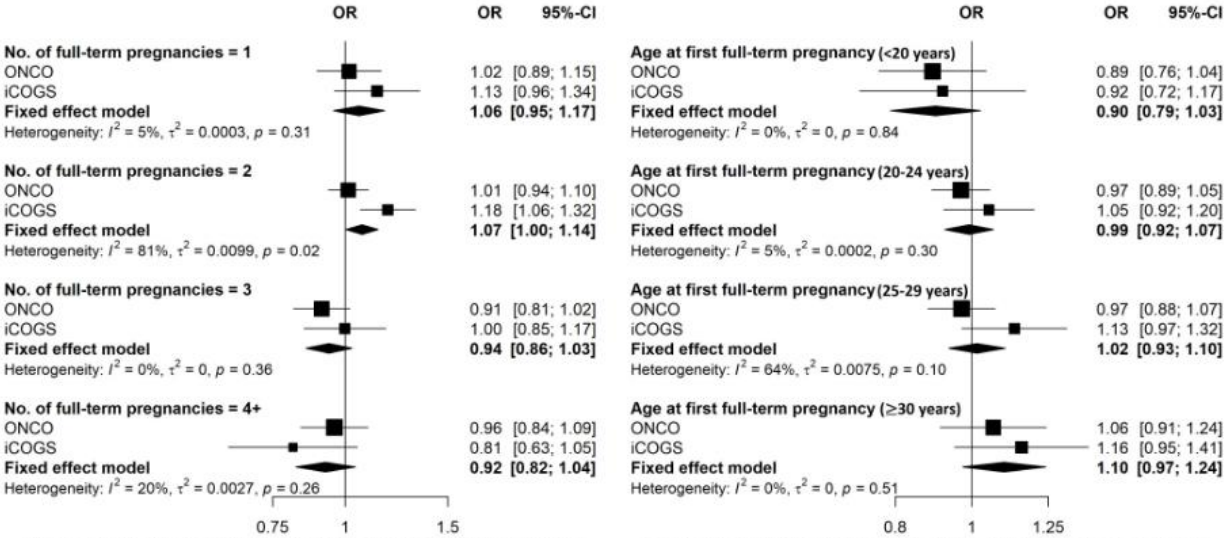
[1] for postmenopausal women only

[2] for parous women only

[3] categories: 1,2,3, ≥4



Models are adjusted for reference age, study, ten principal components, former use of menopausal hormone therapy (MHT), and use of any other type of MHT preparation than the one of interest



Models are adjusted for reference age, study, and ten principal components

Models are adjusted for reference age, study, and ten principal components