IJC
International Journal of Cancer

Gene–environment interactions involving functional variants: Results from the Breast Cancer Association Consortium

Myrto Barrdahl[1], Anja Rudolph[1], John L. Hopper[2], Melissa C. Southey[3], Annegien Broeks[4], Peter A. Fasching[5,6], Matthias W. Beckmann[5], Manuela Gago-Dominguez[7,8], J. Esteban Castelao[9], Pascal Guénel[10], Thérèse Truong[10], Stig E. Bojesen[11,12,13], Susan M. Gapstur[14], Mia M. Gaudet[14], Hermann Brenner[15,16,17], Volker Arndt[15], Hiltrud Brauch[17,18,19], Ute Hamann[20], Arto Mannermaa[21,22,23], Diether Lambrechts[24,25], Lynn Jongen[26], Dieter Flesch-Janys[27,28], Kathrin Thoene[28], Fergus J. Couch[29], Graham G. Giles[2,30], Jacques Simard[31], Mark S. Goldberg[32,33], Jonine Figueroa[34,35], Kyriaki Michailidou[36,37], Manjeet K. Bolla[36], Joe Dennis[36], Qin Wang[36], Ursula Eilber[1], Sabine Behrens[1], Kamila Czene[38], Per Hall[38], Angela Cox[39], Simon Cross[40], Anthony Swerdlow[41,42], Minouk J. Schoemaker[42], Alison M. Dunning[43], Rudolf Kaaks[1], Paul D.P. Pharoah[36,43], Marjanka Schmidt[4,44], Montserrat Garcia-Closas[35], Douglas F. Easton[36,43], Roger L. Milne[2,30] and Jenny Chang-Claude[1,45]

Key words: breast cancer, single nucleotide polymorphism, Breast Cancer Association Consortium, gene–environment, interaction
Additional Supporting Information may be found in the online version of this article.
Conflicts of Interest J.S. is Chairholder of the Canada Research Chair in Oncogenetics. 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. Fergus J. Couch received research support from GRAIL Inc.
Grant sponsor: Cancer Research UK; **Grant number:** C1287/A16563, C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565; **Grant sponsor:** European Union's Horizon 2020 Research and Innovation Programme; **Grant number:** 634935, 633784; **Grant sponsor:** European Community's Seventh Framework Programme; **Grant number:** 223175 (HEALTHF2-2009-223175); **Grant sponsor:** National Institutes of Health; **Grant number:** CA128978; **Grant sponsor:** Post-Cancer GWAS initiative; **Grant number:** 1U19 CA148537, 1U19 CA148065, 1U19 CA148112; **Grant sponsor:** Department of Defence; **Grant number:** W81XWH-10-1-0341; **Grant sponsor:** Canadian Institutes of Health Research (CIHR); **Grant number:** CRN-87521; **Grant sponsor:** Komen Foundation for the Cure; **Grant sponsor:** Breast Cancer Research Foundation; **Grant sponsor:** Ovarian Cancer Research Fund; **Grant sponsor:** United States National Cancer Institute, National Institutes of Health (NIH); **Grant number:** RFA-CA-06-503; **Grant sponsor:** Breast Cancer Family Registry (BCFR), Cancer Care Ontario; **Grant number:** U01 CA69467; **Grant sponsor:** Cancer Prevention Institute of California; **Grant number:** U01 CA69417; **Grant sponsor:** University of Melbourne; **Grant number:** U01 CA69638; **Grant sponsor:** National Cancer Institute; **Grant number:** UM1 CA164920; **Grant sponsor:** National Health and Medical Research Council of Australia; **Grant sponsor:** New South Wales Cancer Council; **Grant sponsor:** Victorian Health Promotion Foundation; **Grant sponsor:** Victorian Breast Cancer Research Consortium; **Grant sponsor:** Dutch Cancer Society; **Grant number:** NKI 2007-3839; 2009-4363; **Grant sponsor:** Dutch Government; **Grant number:** NWO 184.021.007; **Grant sponsor:** Dutch National Genomics Initiative; **Grant sponsor:** ELAN-Fond; **Grant sponsor:** Acción Estratégica de Salud del Instituto de Salud Carlos III; **Grant number:** FIS PI12/02125, PI13/01136; **Grant sponsor:** KAU; **Grant number:** 1-117-1434-HiCi; **Grant sponsor:** Botin Foundation's Fund; **Grant sponsor:** Programa Grupos Emergentes, Cancer Genetics Unit, CHUVI Vigo Hospital, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Instituto de Salud Carlos III, Spain; **Grant sponsor:** 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 number:** 10CSA012E; **Grant sponsor:** Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; **Grant number:** EC11-192; **Grant sponsor:** Grant FEDER-Innterconecta, Ministerio de Economia y Competitividad, Xunta de Galicia, Spain; **Grant sponsor:** Fondation de France; **Grant sponsor:** Institut National du Cancer (INCa); **Grant sponsor:** Ligue Nationale contre le Cancer; **Grant sponsor:** Ligue contre le Cancer Grand Ouest; **Grant sponsor:** Agence Nationale de Sécurité Sanitaire (ANSES); **Grant sponsor:** Agence Nationale de la Recherche (ANR); **Grant sponsor:** Chief Physician Johan Boserup and Lise Boserup Fund; **Grant sponsor:** Danish Medical Research Council; **Grant sponsor:** Herlev Hospital; **Grant sponsor:** American Cancer Society; **Grant sponsor:** Baden Württemberg Ministry of Science, Research and Arts; **Grant sponsor:** German Cancer Aid (Deutsche Krebshilfe); **Grant sponsor:** Federal Ministry of Education and Research (BMBF), Germany; **Grant number:** 01KW9975/5, 01KW9976/8, 01KW9977/0, 01KW0114; **Grant sponsor:** Robert Bosch Foundation, Stuttgart; **Grant sponsor:** Deutsches Krebsforschungszentrum (DKFZ), Heidelberg; **Grant sponsor:** Institute for Prevention and Occupational Medicine of the German Social Accident Insurance; **Grant sponsor:** Institute of the Ruhr University Bochum (IPA), Bochum; **Grant sponsor:** Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany; **Grant sponsor:** Government Funding (EVO) of Kuopio University Hospital; **Grant sponsor:** Cancer Fund of North Savo; **Grant sponsor:** Finnish Cancer Organizations; **Grant sponsor:** University of Eastern Finland; **Grant sponsor:** Stichting tegen Kanker; **Grant number:** 232-2008, 196-2010; **Grant sponsor:** FWO; **Grant number:** KULPFV/10/016-SymBioSysII; **Grant sponsor:** KULPFV; **Grant number:** 70-2892-BR I, 106332, 108253, 108419, 110826, 110828; **Grant sponsor:** Hamburg Cancer Society; **Grant sponsor:** NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer; **Grant number:** CA116201; **Grant sponsor:** David F. and Margaret T.; **Grant sponsor:** Grohne Family Foundation; **Grant sponsor:** VicHealth; **Grant sponsor:** Cancer Council Victoria; **Grant**

[1] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia
[3] Department of Pathology, The University of Melbourne, Melbourne, VIC, Australia
[4] Division of Molecular Pathology, Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
[5] Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany
[6] David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA
[7] Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, Santiago De Compostela, Spain
[8] Moores Cancer Center, University of California San Diego, La Jolla, CA
[9] Oncology and Genetics Unit, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain
[10] CESP–Cancer and Environment team, INSERM U1018, Université Paris-Sud, Université Paris-Saclay, Villejuif, France
[11] Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
[12] Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
[13] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[14] Epidemiology Research Program, American Cancer Society, Atlanta, GA
[15] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[16] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
[17] German Cancer Consortium (DKTK), Heidelberg, Germany
[18] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany
[19] University of Tübingen, Tübingen, Germany
[20] Molecular Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany
[21] Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland
[22] Pathology and Forensic Medicine, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland
[23] Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland
[24] Vesalius Research Center, VIB, Leuven, Belgium
[25] Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium
[26] Leuven Multidisciplinary Breast Center, Department of Oncology, KU Leuven and Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium
[27] Institute for Medical Biometrics and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[28] Department of Cancer Epidemiology, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[29] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN
[30] Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, VIC, Australia
[31] Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, QC, Canada
[32] Department of Medicine, McGill University, Montréal, QC, Canada
[33] Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, Canada
[34] Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh Medical School, Teviot Place Edinburgh, Edinburgh, United Kingdom
[35] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD

sponsor: NHMRC; **Grant number:** 209057, 251553, 504711; **Grant sponsor:** Quebec Breast Cancer Foundation; **Grant sponsor:** CIHR Team in Familial Risks of Breast Cancer; **Grant sponsor:** Ministry of Economic Development, Innovation and Export Trade; **Grant number:** # PSR-SIIRI-701; **Grant sponsor:** Intramural Research Funds, National Cancer Institute, Department of Health and Human Services, USA; **Grant sponsor:** Märit and Hans Rausings Initiative Against Breast Cancer; **Grant sponsor:** Cancer Risk Prediction Center CRisP; **Grant sponsor:** Linnaeus Centre; **Grant sponsor:** Swedish Research Council; **Grant number:** 70867902; **Grant sponsor:** Agency for Science, Technology and Research of Singapore (A*STAR); **Grant sponsor:** US National Institute of Health (NIH); **Grant sponsor:** Susan G. Komen Breast Cancer Foundation; **Grant sponsor:** Yorkshire Cancer Research; **Grant number:** S295, S299, S305PA; **Grant sponsor:** Sheffield Experimental Cancer Medicine Centre; **Grant sponsor:** UK National Institute for Health Research Biomedical Research Centre, University of Cambridge; **Grant sponsor:** Cancer Research UK; **Grant sponsor:** Breast Cancer Now; **Grant sponsor:** Institute of Cancer Research (ICR), London; **Grant sponsor:** Cancer Care Ontario; **Grant number:** U01 CA69467; **Grant sponsor:** National Cancer Institute; **Grant number:** UM1 CA164920; **Grant sponsor:** Acción Estratégica de Salud del Instituto de Salud Carlos III; **Grant number:** FIS PI12/02125 , PI13/01136
DOI: 10.1002/ijc.30859
History: Received 7 Nov 2016; Accepted 11 Apr 2017; Online 3 July 2017

Correspondence to: Jenny Chang-Claude, Im Neuenheimer Feld 581, 69120 Heidelberg, Germany, Tel.: +49-6221-42-2373, Fax: +49-6221-42-2203, E-mail: j.chang-claude@dkfz.de

[36] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Worts Causeway, Cambridge, United Kingdom
[37] Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, Nicosia
[38] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[39] Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, University of Sheffield, Sheffield, United Kingdom
[40] Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, United Kingdom
[41] Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, London, United Kingdom
[42] Division of Breast Cancer Research, The Institute of Cancer Research, Sutton, London, United Kingdom
[43] Department of Oncology, University of Cambridge, Worts Causeway, Centre for Cancer Genetic Epidemiology, Cambridge, United Kingdom
[44] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands
[45] Research Group Genetic Cancer Epidemiology, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Investigating the most likely causal variants identified by fine-mapping analyses may improve the power to detect gene–environment interactions. We assessed the interplay between 70 single nucleotide polymorphisms identified by genetic fine-scale mapping of susceptibility loci and 11 epidemiological breast cancer risk factors in relation to breast cancer. Analyses were conducted on up to 58,573 subjects (26,968 cases and 31,605 controls) from the Breast Cancer Association Consortium, in one of the largest studies of its kind. Analyses were carried out separately for estrogen receptor (ER) positive (ER+) and ER negative (ER–) disease. The Bayesian False Discovery Probability (BFDP) was computed to assess the noteworthiness of the results. Four potential gene–environment interactions were identified as noteworthy (BFDP < 0.80) when assuming a true prior interaction probability of 0.01. The strongest interaction result in relation to overall breast cancer risk was found between *CFLAR*-rs7558475 and current smoking ($OR_{int} = 0.77$, 95% CI: 0.67–0.88, $p_{int} = 1.8 \times 10^{-4}$). The interaction with the strongest statistical evidence was found between 5q14-rs7707921 and alcohol consumption ($OR_{int} = 1.36$, 95% CI: 1.16–1.59, $p_{int} = 1.9 \times 10^{-5}$) in relation to ER– disease risk. The remaining two gene–environment interactions were also identified in relation to ER– breast cancer risk and were found between 3p21-rs6796502 and age at menarche ($OR_{int} = 1.26$, 95% CI: 1.12–1.43, $p_{int} = 1.8 \times 10^{-4}$) and between 8q23-rs13267382 and age at first full-term pregnancy ($OR_{int} = 0.89$, 95% CI: 0.83–0.95, $p_{int} = 5.2 \times 10^{-4}$). While these results do not suggest any strong gene–environment interactions, our results may still be useful to inform experimental studies. These may in turn, shed light on the potential interactions observed.

What's new?

Although it is widely acknowledged that genes and environment may interact to cooperatively modify breast cancer risk, no such interaction is known at the single nucleotide polymorphism (SNP) level. Here, the authors assessed the interplay of 70 SNPs with 11 known breast cancer risk factors in estrogen receptor-positive and -negative disease. Weak interactions were found with individual SNPs and current smoking or alcohol consumption but no strong gene–environment interaction was identified. These data do not support the model of strong modification of genetic cancer risk by environmental factors.

In 1968, MacMahon stated that "In no field are there more complex examples of the gene–environment relationship than in experimental cancer research."[1]. Following his words and the general opinion that genetic and non-genetic risk factors do not give rise to the disease solely by acting on independent pathways, several studies have investigated gene–environment interplay in relation to breast cancer risk. Studies of this type are motivated by the fact that the identification of gene–environment interactions in relation to breast cancer could provide insight into the biological mechanisms underlying the disease, allow the distinction of women at high risk from women at lower risk and improve the accuracy of risk prediction models. However, despite large-scale, international efforts, to date, there are few single nucleotide polymorphisms (SNPs) for which the effect on breast carcinogenesis has been found to be modified by an epidemiological risk factor, and only one of these has been replicated.[2,3]

Several breast cancer risk *loci* that were previously identified in genome-wide association studies (GWAS) were recently investigated further by genetic fine-scale mapping in the framework of the Collaborative Oncological Gene-Environment Study (COGS) using samples from studies participating in the Breast Cancer Association Consortium (BCAC). The SNPs identified in the fine-mapping studies were further investigated in subsequent functional studies to identify potential causal associations. The consideration of causal variants may improve power to detect gene–environment interplay. However, if no interactions are detected, the weight of evidence against gene–environment interactions for the *locus* in question is strengthened. Additionally, new susceptibility alleles were

identified from genotypes generated by imputation using the 1000 Genomes Project reference panel. Therefore, in this analyses, multiplicative gene–environment interaction in relation to breast cancer risk was assessed between 55 potentially causal as well as 15 newly identified SNP alleles, and the following 11 established epidemiological risk factors: age at menarche, oral contraceptive (OC) use, parity, age at first full-term pregnancy (FTP), number of FTPs, breastfeeding, use of menopausal hormone therapy (MHT), body mass index (BMI), adult height, smoking and alcohol consumption. We also investigated interactions in relation to estrogen receptor (ER) specific breast cancer risk since different disease subtypes may arise through different pathways. The analyses reported in this article are based on the largest, currently available dataset with genetic data and extensive epidemiological information.

Methods

Study subjects

Data on subjects of European descent derived from 21 studies participating in the BCAC were pooled. A brief description of each study can be found in Supporting Information Table S1. There were 12 population-based and 9 non-population based studies, each contributing at least 200 cases and 200 controls with available SNP data and information on at least one epidemiological risk factor. Subjects were excluded from the gene–environment interaction analyses if they were male, of non-European origin, a prevalent case or had missing data on age at diagnosis or age at interview, the epidemiological risk factor in question or any of the adjustment variables. Hence, the number of study subjects for each SNP-risk factor pair varied with the availability of epidemiological data. Analyses were based on between 11,342 subjects (5,445 cases and 5,897 controls) for effect modification by alcohol consumption and 58,573 subjects (26,968 cases and 31,605 controls) for effect modification by number of FTPs. The set of study subjects that were included in at least one gene–environment interaction model comprised 30,000 cases and 34,501 controls. All studies were approved by the relevant ethics committees and informed consent was obtained from all participants.

SNP selection and genotyping

Genotyping was carried out using an Illumina iSelect array (iCOGS) in the framework of the COGS project (www.nature.com/icogs). With the aim of detecting causal variants, a number of *loci* known to confer breast cancer risk at the time of the design of the iCOGS array were further investigated using fine scale genetic mapping. To improve SNP density, imputation of the respective regions was performed using the March 2012 release of the 1000 Genomes as reference panel. The functional follow-up work was not carried out centrally for all regions but divided between the different working groups of BCAC and thus the methods used varied somewhat.[4–17] In addition, imputed genotypes for 15 new susceptibility loci identified through a meta-analysis of 11 GWAS with genotypes SNPs generated by imputation using the 1000 Genomes Project March 2012 release as the reference panel were used.[5] A list of the 70 SNPs included in the analyses for this report can be found in Supporting Information Table S2.

Data filtering

Data from the participating studies were centrally cleaned and harmonized. The information on epidemiological factors was collected at date of reference. In the case–control studies, this was defined as the date of diagnosis for cases and the date of questionnaire for controls, and in the three cohort studies (Cancer Prevention Study II [CPSII], Melbourne Collaborative Cohort Study [MCCS] and UK Breakthrough Generations Study [UKBGS]) information at baseline was used, unless follow-up information was available. Women who were 54 years or younger at reference were considered pre-menopausal and women who were older than 54 years at reference were considered postmenopausal. Subjects who were smokers within 1 year before reference date or used MHT within 6 months before reference date were considered to be current smokers and current MHT users. For the case–control studies, BMI was calculated using usual adult weight or weight 1 year before reference (Australian Breast Cancer Family Study, CECILE Breast cancer Study, Gene-Environment Interaction and Breast Cancer in Germany, kConFab, Kuopio Breast Cancer Project, Mammary Carcinoma Risk Factor Investigation, Mayo Clinic Breast Cancer Study, NCI Polish Breast Cancer Study and Singapore and Sweden Breast Cancer Study), or weight around the age of 20 years (ESTHER Breast Cancer Study, Karolinska Mammography Project for Risk Prediction of Breast Cancer-prevalent cases and Study of Epidemiology and Risk Factors in Cancer Heredity). For the cohort studies (CPSII, MCCS and UKBGS), BMI was calculated using information from baseline or the latest available questionnaire before diagnosis, if available.

Statistical analysis

Association analyses of SNP alleles and breast cancer risk were carried out using logistic regression models adjusted for age at reference, study and ethnicity. In all models used in this study, genotyped SNPs were treated as ordinal variables (counts of minor alleles) and imputed SNPs as continuous variables.

The main effects of the epidemiological risk factors were also investigated using logistic regression models adjusted for reference age, study and self-assessed ethnicity. Heterogeneity across studies was explored by means of Cochrane's Q-test. The epidemiological variables used in this analyses were categorized as follows: age at menarche (per 2 years), ever use of OC (yes or no), ever parous (yes or no), number of FTPs for parous women (1, 2, 3 and $\geq$4 FTPs), ever breastfed (yes or no), age at first FTP (per 5 years), adult BMI for pre- and postmenopausal women, respectively (per 5 kg/m^2), adult height (per 5 cm), current use of MHT in the form of estrogen and progesterone or estrogen only (yes or no), lifetime

average alcohol intake (per 10 g/day), current smoking (yes or no) and pack-years of smoking (per 10 pack-years).

Multiplicative gene–environment interaction was assessed by comparing logistic regression models with and without SNP-risk factor interaction terms by means of the likelihood ratio test. All models on which this study is based were adjusted for study, reference age and ethnicity, so as to capture genetic population sub-structure. An interaction term between the epidemiological variable and an indicator for population based study design was included to protect against bias due to the differing selection of study participants in non-population based *versus* population-based studies. Interactions of SNPs and epidemiological risk factors were also investigated in relation to ER specific (ER+ or ER–) breast cancer risk, using cases and controls. Furthermore, potentially differential gene–environment interaction according to ER status was assessed in case-only analyses comparing ER– cases to ER+ cases. The ER-specific models and the case-only analyses were adjusted similarly as the interaction models for overall breast cancer risk. To elucidate the results of the interaction analyses, risk association between SNPs and breast cancer was investigated by stratifying on the epidemiological variables.

MHT was sub-divided into estrogen only and combined (estrogen plus progestogen) therapy and investigated in relation to breast cancer risk using only post-menopausal women. All statistical models involving MHT use were further adjusted for former MHT use and current use of the MHT preparation (estrogen only or combined) not included in the interaction term. Additionally, interactions of SNPs and BMI for postmenopausal women were assessed in never- and former users of MHT only. All risk analyses were carried out using SAS 9.2.

Between-study heterogeneity of the interaction odds ratio (OR) estimates was investigated using Cochrane's Q-test and quantified by the ratio of true heterogeneity to the total observed variation, denoted I^2. Heterogeneity was investigated for SNP-risk factor pairs with an interaction p values below the Bonferroni corrected threshold of statistical significance for genetic main effects, computed by dividing the standard threshold of 0.05 by the number of SNPs ($0.05/70 > 7 \times 10^{-4}$). Interaction ORs were tested for heterogeneity across studies on basis of interaction p values in models of overall or ER specific breast cancer risk, although the latter on the condition that a heterogeneity $p < 0.05$ of ER+ *versus* ER– disease had been observed. Heterogeneity tests were conducted using the R package "rmeta" (version 2.2).

The Bayesian False Discovery Probability (BFDP) was computed to control the number of false-positive findings and assess the noteworthiness of the results.[18] The cut-off for noteworthiness is based on the ratio of the cost of a false non-discovery to the cost of a false discovery. As suggested in the literature, we set the cost of failing to discover a true association to four times the cost of a falsely reported one, classifying results with a BFDP < 0.8 as noteworthy. The BFDP was calculated for all SNP-risk factor pairs with an interaction p values below the Bonferroni-corrected threshold given above ($p < 7 \times 10^{-4}$). The BFDP was computed for two different prior probabilities of this (0.01, 0.001), under the assumption that the probability of observing a true interaction OR inside the interval 0.66–1.5 was 95%. As a complementary measure to the BFDP, we also computed the ABF, which approximates the ratio of the probability of the data given that the null hypothesis is true, to the probability of the data given that the alternative hypothesis is true. The null hypothesis in this case is that the coefficient of the interaction term in the logistic regression model is equal to zero.

Results

The studies included in the gene–environment interaction analyses are listed in Table 1 together with the number of cases and controls, overall and by ER status. The median time between questionnaire and diagnosis was 3 years in the MCCS cohort, 2 years in the UKBGS cohort and 7 years in the CPSII cohort.

The associations between SNP alleles and breast cancer risk in the subset of BCAC studies with risk factor data available were consistent with earlier reports and can be found in Supporting Information Table S3.[4–14]

Main effects of the epidemiological variables on breast cancer risk across studies are presented in Supporting Information Figure 1. These analyses were carried out using only population-based studies and the results were consistent with what has been reported earlier in the literature.[3,19–30] Current use of OC, MHT use (E only, as well as E + P), alcohol consumption, height as well as never having breastfed (*vs.* ever having breastfed) were all factors that showed an increased risk of breast cancer. A reduction in risk was observed for older age at menarche, ever being parous, number of FTPs and high BMI for pre-menopausal women. For current smoking and pack-years of smoking, no significant association with breast cancer risk was detected.

The complete results from the interaction analyses, showing the risk association between SNPs and breast cancer across categories of the epidemiological variables, are presented in Supporting Information Table S4. We identified four SNP-risk factor pairs with at least one interaction $p < 7 \times 10^{-4}$ in relation to overall, ER+ or ER– breast cancer risk, as presented in Table 2. All of these interactions were classified as noteworthy (BFDP < 0.8) assuming a prior probability of true interaction of 0.01 but no result remained noteworthy at the 0.001 level (Table 3).

First evidence of an interaction in relation to overall disease risk was noted between the variant *CFLAR*-rs7558475 and current smoking ($OR_{int} = 0.77$, 95% CI: 0.67–0.88, $p_{int} = 1.8 \times 10^{-4}$). This result was considered noteworthy (BFDP = 0.40) assuming a prior probability of true interaction of 0.01 and the approximated Bayes factor (ABF) = 0.007 indicated that the data were almost 140 times more likely given the alternative hypothesis than given the null. Breast cancer risk was reduced for current smokers carrying the minor allele (G) ($OR_{per\text{-}allele} = 0.76$, 95% CI: 0.66–

Table 1. Participating studies

Study	Full study name	Study design	Country	Cases			Controls
				All	ER–	ER+	
ABCFS	Australian Breast Cancer Family Study	Population-based	Australia	790	261	456	551
ABCS	Amsterdam Breast Cancer Study	Mixed	Netherlands	1,245	292	800	1,177
BBCC	Bavarian Breast Cancer Cases and Controls	Mixed	Germany	553	86	456	457
BREOGAN	Breast Oncology Galicia Network	Mixed	Spain	1,561	329	1,251	1,423
CECILE	CECILE Breast cancer Study	Population-based	France	900	128	751	999
CGPS	Copenhagen General Population Study	Mixed	Denmark	2,209	269	1,592	4,506
CPSII	Cancer Prevention Study II	Population-based	USA	1,655	35	1,205	1,940
ESTHER	ESTHER Breast Cancer Study	Population-based	Germany	471	98	302	502
GENICA	Gene-Environment Interaction and Breast Cancer in Germany	Population-based	Germany	456	114	333	427
KBCP	Kuopio Breast Cancer Project	Population-based	Finland	411	93	303	251
LMBC	Leuven Multidisciplinary Breast Centre	Mixed	Belgium	2,424	378	2,069	1,045
MARIE	Mammary Carcinoma Risk Factor Investigation	Population-based	Germany	1,656	371	1,278	1,778
MCBCS	Mayo Clinic Breast Cancer Study	Mixed	USA	1,554	254	1,295	1,893
MCCS	Melbourne Collaborative Cohort Study	Population-based	Australia	478	117	343	490
MTLGEBCS	Montreal Gene-Environment Breast Cancer Study	Population-based	Canada	489	64	421	436
PBCS	NCI Polish Breast Cancer Study	Population-based	Poland	519		519	424
pKARMA	Karolinska Mammography Project for Risk Prediction of Breast Cancer-prevalent cases	Mixed	Sweden	2,822	410	2,328	5,469
SASBAC	Singapore and Sweden Breast Cancer Study	Population-based	Sweden	1,163	144	663	1,378
SBCS	Sheffield Breast Cancer Study	Mixed	UK	751	107	367	848
SEARCH	Study of Epidemiology and Risk Factors in Cancer Heredity	Mixed	UK	7,478	1,119	5,371	8,050
UKBGS	UK Breakthrough Generations Study	Population-based	UK	415	47	231	457
Total				30,000	4,716	22,334	34,501

0.88, $p = 2.2 \times 10^{-4}$) compared to that of non-smoker carriers ($OR_{per\text{-}allele} = 0.99$, 95% CI: 0.91–1.08, $p = 0.9$) where no risk association was observed. When comparing ER– cases to ER+ cases, the results did not indicate any effect heterogeneity ($p_{het} = 0.48$). There was no strong evidence of interaction, neither with respect to ER+ risk ($p_{int} = 0.0014$) nor with respect to ER– risk ($p_{int} = 0.75$).

The most promising result of the gene–environment interaction analyses in terms of noteworthiness was considered noteworthy at the 0.01 probability level and was noted between the variant 5q14-rs7707921 located in an intron of the autophagy related 10 (*ATG10*) gene, and alcohol consumption ($OR_{int} = 1.36$, 95% CI: 1.16–1.59, $p_{int} = 1.9 \times 10^{-5}$) in relation to ER– breast cancer. This result had the lowest BFDP = 0.33, and conditioning on the alternative, the data were about 200 times more likely as compared to conditioning on the null (ABF = 0.005). Carriers of the minor allele (T) of rs7707921 had an increased risk of ER– breast cancer if they consumed >20 g of alcohol per day ($OR_{per\text{-}allele} = 2.56$, 95% CI: 1.45–4.62, $p = 0.001$), but not if they consumed <20 g of alcohol per day ($OR_{per\text{-}allele} = 1.07$, 95% CI: 0.92–1.24, $p = 0.36$). A strong effect heterogeneity was detected when comparing ER– cases to ER+ cases ($p_{het} = 6.7 \times 10^{-6}$). Together with the absence of interaction in relation to ER+ disease ($p_{int} = 0.79$) and overall breast cancer risk ($p_{int} = 0.70$), this indicated that the interaction might be specific to ER– disease.

In addition, indications of two further interactions were noted in relation to ER– disease risk. One of these was between 3p21-rs6796502 and age at menarche ($OR_{int} = 1.26$, 95% CI: 1.12–1.43, $p_{int} = 1.8 \times 10^{-4}$) which had BFDP = 0.49, and of which the ABF (ABF = 0.010) implied that the data were 100 times more likely under the alternative hypothesis than under the null. Carriers of the minor allele (A) of 3p21-rs6796502 who experienced their menarche no later than the age of 11 years had a reduced risk of ER– breast cancer ($OR_{per\text{-}allele} = 0.70$, 95% CI: 0.54–0.90, $p = 0.006$), whereas there was no association with disease risk of the genetic variant for women who had their menarche between the age of 12 and 13 years ($OR_{per\text{-}allele} = 0.88$,

95% CI: 0.76–1.02, $p = 0.08$), or after the age of 14 years ($OR_{per\text{-}allele} = 1.16$, 95% CI: 0.99–1.34, $p = 0.06$). While the observed interaction was in relation to ER– disease risk, no effect heterogeneity was detected when comparing ER– and ER+ cases ($p_{het} = 0.53$) nor was there any indication of any interaction in relation to overall breast cancer risk ($p_{int} = 0.94$). Hence, it is not possible to conclude that the observed interaction is specific for ER– disease.

Finally, an indication of a gene–environment interaction was found between 8q23-rs13267382 and age at first FTP ($OR_{int} = 0.89$, 95% CI: 0.83–0.95, $p_{int} = 5.2 \times 10^{-4}$) in relation to ER– disease risk. This interaction had BFDP = 0.61 assuming a true, prior interaction probability of 0.01, and ABF = 0.016, indicating that the data are about 60 times more likely conditioning on the alternative than on the null. There was no interaction observed in relation to disease risk, when considering ER+ breast cancer ($p_{int} = 0.98$), or overall breast disease risk ($p_{int} = 0.47$), and no effect heterogeneity was found when comparing the risk of ER– and ER+ breast cancer ($p_{het} = 0.99$). Our findings indicated a modest reduction in ER– breast cancer risk for minor allele (A) carriers who were aged 30 or above at their first FTP ($OR_{per\text{-}allele} = 0.79$, 95% CI: 0.68–0.91, $p = 0.001$), whereas for women who had their first child at younger ages the allele had no effect on risk.

Discussion

From the analyses presented in this work, four SNP-risk factor pairs were identified, for which $p_{int} < 7 \times 10^{-4}$, and all of the interactions were considered noteworthy (BFDP < 0.8) assuming a true prior interaction probability of 0.01. One of the results was detected in relation to overall breast cancer risk, while the three remaining results appeared to be specific for ER– disease.

The strongest gene–environment interaction in relation to overall breast cancer risk was noted between rs7558475 located in the *CASP8* and *FADD* like apoptosis regulator (*CFLAR*) gene and current smoking ($p_{int} = 1.8 \times 10^{-4}$). The protein product of *CFLAR* regulates apoptosis, thus it is possible that *CFLAR* genetic variants affect response to DNA damage caused by tobacco associated carcinogens and therefore modify breast cancer risk conferred by smoking. However, although rs7558475 is located in a *CFLAR* enhancer region, reports from recent functional studies and expression quantitative trait locus (eQTL) analyses did not provide any convincing evidence regarding functionality.[6,31] Hence, further work is required to understand possible biological mechanism related to the observed interaction.

The strongest statistical evidence of interaction was found in relation to ER– breast cancer risk and was noted between an intron variant 5q14-rs7707921 in the autophagy related 10 (*ATG10*) gene, and alcohol consumption ($p_{int} = 1.9 \times 10^{-5}$). Autophagy, which is considered a survival mechanism of the cell, may act as a tumor suppressor but also influence cell survival by promoting tumor growth, and has been suggested as a target in cancer therapy.[32] It has been reported that autophagy could have a protective effect on esophageal epithelial cells responding to ethanol-induced oxidative stress.[33] Also, while ethanol promotes oxidative stress in cancer associated fibroblasts, it has been reported to induce autophagy resistance in epithelial cells.[34] Given the above information, it is conceivable that alcohol consumption could influence the effect on breast cancer risk of an autophagy-related polymorphism. However the biological mechanism needs to be further investigated. The position of the variant *ATG10*-rs7707921 does not coincide with any strong regulatory elements. The eQTL analyses carried out within the framework of BCAC showed a strong association between the T allele of rs7707921 and expression of the ribosomal protein S23 gene (*RPS23*) in breast tissue as well as a moderate association between the allele and expression of the ATPase, H+ transporting, lysosomal accessory protein 1-like (*ATP6AP1L*) gene.[5] The *RPS23* gene encodes a ribosomal protein and the *ATP6AP1L* is also protein coding but the genes have not yet been implicated in ER– breast cancer risk and their expression levels have not been assessed in relation to alcohol consumption or oxidative stress. Further work is thus needed to understand how the protein products of these genes could interact with alcohol consumption to modify the risk association of rs7707921 with ER– breast cancer.

Furthermore, we found an indication of a possible interaction between 3p21-rs6796502 and age at menarche ($p_{int} = 1.8 \times 10^{-4}$) in relation to ER– breast cancer. Our results suggest that the reduced risk of ER– breast tumors for carriers of the A-allele are modified for women with late age at menarche ≥ 14 years. However, according to a recent functional study, the SNP is not located in the vicinity of any genes or enhancer regions in mammary cell lines, nor are there any significant results available from eQTL analyses.[5] In addition, no significant effect heterogeneity was found when comparing the interaction between ER– and ER+ cases to support that the result could to be specific to ER– disease. It is thus necessary to first confirm this interaction with further data before attempting any biological explanation.

The interaction observed between the intron variant 8q23-rs13267382 of the long intergenic non protein coding RNA 536 gene (*LINC00536*) and age at first FTP ($p_{int} = 2.6 \times 10^{-4}$) suggests that the variant is associated with a reduced risk of ER– breast cancer with older age at first FTP, whereby the association was statistically significant for women who were at least 30 years of age at their first FTP. Overall, this variant was not reported to be associated with ER– disease risk,[5] which is confirmed in the current report. Neither this SNP nor any variants in high linkage disequilibrium with it are positioned in the vicinity of any regulatory genomic feature. As for the interaction with 3p21-rs6796502, there was not clear evidence for this interaction to be specific to ER– disease. Therefore, further data are required to confirm this interaction.

Table 2. SNP-risk factor pairs with interaction $p < 7 \times 10^{-4}$, overall and by ER status across categories of epidemiological risk factors

			Overall		ER+		ER−		
SNP/risk factor	Stratum	Cases/controls	OR (95% CI)	$p_{interaction}$	OR (95% CI)	$p_{interaction}$	OR (95% CI)	$p_{interaction}$	$P_{case\text{-}only}$
CFLAR-rs7558475/ current smoking	No	7,698/8,835	0.99 (0.91–1.08)		1.00 (0.91–1.09)		0.96 (0.81–1.14)		
	Yes	1,375/1,519	0.76 (0.66–0.88)		0.78 (0.66–0.91)		0.91 (0.63–1.31)		
	All	9,073/10,354	0.96 (0.88–1.04)	1.8×10^{-4}	0.96 (0.88–1.05)	0.0014	0.96 (0.82–1.12)	0.75	0.48
5q14-rs7707921/ alcohol consumption	<20 g/day	4,904/5,411	1.16 (1.08–1.24)		1.18 (1.10–1.27)		1.07 (0.92–1.24)		
	≥20 g/day	481/541	1.16 (0.94–1.43)		1.08 (0.86–1.36)		2.59 (1.45–4.62)		
	All	5,385/5,952	1.16 (1.09–1.23)	0.70	1.17 (1.09–1.25)	0.79	1.15 (0.99–1.32)	1.9×10^{-5}	6.7×10^{-6}
3p21-rs6796502/ age at menarche	≤11 years	3,350/3,609	0.73 (0.65–0.83)		0.75 (0.65–0.86)		0.70 (0.54–0.90)		
	12–13 years	9,503/10,893	0.93 (0.86–0.99)		0.93 (0.86–1.00)		0.88 (0.76–1.02)		
	≥14 years	7,294/8,864	0.95 (0.88–1.03)		0.92 (0.84–1.01)		1.16 (0.99–1.34)		
	All	20,147/23,366	0.90 (0.86–0.95)	0.94	0.89 (0.85–0.94)	0.73	0.94 (0.86–1.04)	1.8×10^{-4}	0.53
8q23-rs13267382/ age at first FTP	<20 years	2,085/1,830	1.01 (0.92–1.11)		1.00 (0.90–1.11)		1.12 (0.94–1.33)		
	20–24 years	6,944/8,246	0.92 (0.88–0.97)		0.91 (0.86–0.96)		1.00 (0.91–1.11)		
	25–29 years	5,388/6,877	0.97 (0.92–1.02)		0.98 (0.92–1.03)		0.91 (0.81–1.02)		
	≥30 years	2,965/3,555	0.92 (0.85–0.99)		0.94 (0.87–1.02)		0.79 (0.68–0.91)		
	All	17,382/20,508	0.94 (0.92–0.97)	0.47	0.95 (0.91–0.98)	0.98	0.95 (0.89–1.01)	5.2×10^{-4}	0.99

Table 3. BFDPs for SNP-risk factor pairs with interaction $p < 7 \times 10^{-4}$

Breast cancer subtype	SNP/risk factor	$OR_{interaction}$ (95%CI)	BFDP[1], prior probability of true interaction		ABF[2]
			0.01	0.001	
Overall	*CFLAR*-rs7558475/current smoking	0.77 (0.67–0.88)	0.40	0.87	0.007
ER–	5q14-rs7707921/alcohol	1.36 (1.16–1.59)	0.33	0.83	0.005
ER–	3p21-rs6796502/age at menarche	1.26 (1.12–1.43)	0.49	0.91	0.010
ER–	8q23-rs13267382/age at first FTP	0.89 (0.83–0.95)	0.61	0.94	0.016

[1]The BFDP was calculated assuming that the true interaction OR is between 0.66 and 1.50. [2]ABF is an approximation of the rate of the probability of the data given the null to the probability of the data given the alternative hypothesis.

This work is subject to a number of limitations. First, despite central harmonization of the data, substantial heterogeneity was observed in the risk estimates of the epidemiological risk factors across studies, which brought about the inclusion of a product term of study design and epidemiological variable in the interaction models, and the quantification of epidemiological main effects based on the population based studies. Second, the study population consisted predominantly of case-control studies (only three cohort studies), which are known to be prone to selection bias and recall bias, as well as associated misclassification of risk factors. However, gene–environment interaction estimates were similar in the overall study population compared to the subset of population based studies (data not shown). Misclassification of epidemiological risk factors is known to reduce the power to detect interactions, rather than increasing the probability of false-positive findings.[35] Hence, this study is more likely to be subject to limited power than to spurious gene–environment interactions. Also, our findings are based on study participants of Caucasian origin so that they may not be generalizable to other populations. For the ER specific risk analyses, in particular in the subgroup of ER– cases ($N = 4{,}662$), the power was diminished due to the reduced sample size.

However, this study also has several strengths. To begin with, the interaction analyses are based on the largest dataset presently available. The four indicated interactions were based on 11,337 subjects (5,385 cases and 5,952 controls) in analyses with respect to alcohol consumption, and 19,427 subjects (9,073 cases and 10,354 controls) for current smoking, as well as 43,513 subjects (20,147 cases and 23,366 controls) in the analyses of age at menarche and 37,819 subjects (17,382 cases and 20,508 controls) in the analyses of age at first FTP.

Taken together, the results presented in this report are not in line with the existence of strong modification of the allelic effects on breast cancer risk by the epidemiological risk factors investigated. However, the results presented in this report contribute to the global body of knowledge on gene–environment interactions by generating hypotheses, thereby providing guidance for future functional studies and large scale replication studies.

Acknowledgements

The authors thank the contributions of Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), 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, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie La Boissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, 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. Australian Breast Cancer Family Study thanks Maggie Angelakos, Judi Maskiell and Gillian Dite. Amsterdam Breast Cancer Study thanks Sanquin bloodbank, The Netherlands, Sten Cornelissen, Richard van Hien, Linde Braaf, Frans Hogervorst, Senno Verhoef, Laura van 't Veer, Emiel Rutgers, C Ellen van der Schoot and Femke Atsma. Breast Oncology Galicia Network thanks Angel Carracedo, Maria Elena Martinez, Victor Muñoz Garzón, Alejandro Novo Domínguez, Sara Miranda Ponte, Carmen M Redondo, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza, Francisco Gude Sampedro, 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 Investigación Sanitaria Galicia Sur (IISGS), SERGAS, Vigo, Spain. Copenhagen General Population Study thanks Staff and participants of the Copenhagen General Population Study. The author also thank the following for the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank and Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. ESTHER Breast Cancer Study thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier and Katja Butterbach. The Gene-Environment Interaction and Breast Cancer in Germany network: Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [HB, Wing-Yee Lo, Christina Justenhoven], German Cancer Consortium (DKTK) and German Cancer Research Center (Deutsches Krebsforschungszentrum [DKFZ]) [HB], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [Yon-Dschun Ko, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, 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]. Kuopio Breast Cancer Project thanks Eija Myöhänen and Helena Kemiläinen. Leuven Multidisciplinary Breast Centre thanks Gilian Peuteman, Dominiek Smeets, Thomas

Van Brussel and Kathleen Corthouts. Mammary Carcinoma Risk Factor Investigation thanks Petra Seibold, Judith Heinz, Nadia Obi, Alina Vrieling, Muhabbet Celik, Til Olchers and Stefan Nickels. Montreal Gene-Environment Breast Cancer Study thanks Martine Tranchant (CHU de Québec Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management and skillful technical assistance. NCI Polish Breast Cancer Study thanks: Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao and Michael Stagner. Karolinska Mammography Project for Risk Prediction of Breast Cancer-prevalent cases and Singapore and Sweden Breast Cancer Study thank The Swedish Medical Research Counsel and Märit and Hans Rausings Initiative Against Breast Cancer. Sheffield Breast Cancer Study thanks Sue Higham, Helen Cramp, Ian Brock, Sabapathy Balasubramanian and Dan Connley. Study of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) thanks The SEARCH and EPIC teams. UK Breakthrough Generations Study 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. This work was supported by NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre a well as multiple funding agencies: funding for BCAC and the iCOGS infrastructure came from: Cancer Research UK (C1287/A16563, C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST, respectively), the European Community's Seventh Framework Programme under grant agreement no. 223175 (HEALTHF2-2009-223175) (COGS), the National Institutes of Health (NIH) (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, Komen Foundation for the Cure, the Breast Cancer Research Foundation and the Ovarian Cancer Research Fund. The ABCFS work was supported by the United States National Cancer Institute, NIH, under RFA-CA-06-503, and through cooperative agreements with members of the Breast Cancer Family Registry (BCFR) and Principal Investigators, including Cancer Care Ontario (U01 CA69467), Cancer Prevention Institute of California (U01 CA69417) and University of Melbourne (U01 CA69638). The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA), 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. The Amsterdam Breast Cancer Study was supported by the Dutch Cancer Society (grants NKI 2007-3839; 2009-4363); BBMRI-NL, which is a Research Infrastructure financed by the Dutch government (NWO 184.021.007); and the Dutch National Genomics Initiative. The work of the Bavarian Breast Cancer Cases and Controls was partly funded by ELAN-Fond of the University Hospital of Erlangen. Breast Oncology Galicia Network is funded by FIS PI12/02125 and PI13/01136 of Acción Estratégica de Salud del Instituto de Salud Carlos III; KAU grant no. (1-117-1434-HiCi); the Botin Foundation's Fund; Programa Grupos Emergentes, Cancer Genetics Unit, CHUVI Vigo Hospital, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Instituto de Salud Carlos III, Spain; 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 number 10CSA012E); Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain (grant number EC11-192) and Grant FEDER-Innterconecta, Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. The CECILE Breast cancer Study was funded by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Ligue contre le Cancer Grand Ouest, Agence Nationale de Sécurité Sanitaire (ANSES), Agence Nationale de la Recherche (ANR). The Copenhagen General Population Study was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital. The recruitment and maintenance of CPS-II is supported by the American Cancer Society. The ESTHER Breast Cancer 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 Gene-Environment Interaction and Breast Cancer in Germany 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 Kuopio Breast Cancer Project was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations and by the strategic funding of the University of Eastern Finland. Leuven Multidisciplinary Breast Centre is supported by the Stichting tegen Kanker (232-2008 and 196-2010) and by the FWO and the KULPFV/10/016-SymBioSysII (to D.L.). The Mammary Carcinoma Risk Factor Investigation study was supported by the Deutsche Krebshilfe e.V. (70-2892-BR I, 106332, 108253, 108419, 110826 and 110828), the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF), Germany (01KH0402). The Mayo Clinic Breast Cancer Study 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. Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria, by Australian NHMRC grants 209057, 251553 and 504711 and 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 work of Montreal Gene-Environment Breast Cancer Study 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 (CRN-87521) and the Ministry of Economic Development, Innovation and Export Trade grant (PSR-SIIRI-701). The NCI Polish Breast Cancer Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Karolinska Mammography Project for Risk Prediction of Breast Cancer-prevalent cases is a combination of the KARMA and LIBRO-1 studies. KARMA was supported by Märit and Hans Rausings Initiative Against Breast Cancer. KARMA and LIBRO-1 were supported by the Cancer Risk Prediction Center (CRisP; www.crispcenter.org), a Linnaeus Centre (Contract ID 70867902)] financed by the Swedish Research Council. The Singapore and Sweden Breast Cancer 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 Sheffield Breast Cancer Study was supported by Yorkshire Cancer Research (S295, S299 and S305PA) and Sheffield Experimental Cancer Medicine Centre. Study of Epidemiology and Risk Factors in Cancer Heredity is funded by a programme grant from Cancer Research UK (C490/A10124) and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The UK Breakthrough Generations Study is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.

References

1. MacMahon B. Gene-environment interactions in human disease. *J PsychiatrRes* 1968;6:393–402.
2. Rudolph A, Chang-Claude J, Schmidt MK. Gene-environment interaction and risk of breast cancer. *Br J Cancer* 2016;114:125–33.
3. Milne RL, Gaudet MM, Spurdle AB, et al. Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: a combined case-control study. *Breast Cancer Res* 2010;12:R110.
4. Darabi H, McCue K, Beesley J, et al. Polymorphisms in a putative enhancer at the 10q21.2 breast cancer risk locus regulate NRBF2 expression. *Am J Hum Genet* 2015;97:22–34.
5. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015;47:373–80.
6. Lin WY, Camp NJ, Ghoussaini M, et al. Identification and characterization of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. *Hum Mol Genet* 2015; 24:285–98.
7. Ghoussaini M, Edwards SL, Michailidou K, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun* 2014;4:4999
8. Bojesen SE, Pooley KA, Johnatty SE, 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: 371–84. 84e1–2.
9. Milne RL, Burwinkel B, Michailidou K, et al. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* 2014;23:6096–111.
10. Meyer KB, O'reilly M, Michailidou K, 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; 93:1046–60.
11. French JD, Ghoussaini M, Edwards SL, 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;92:489–503.
12. Guo X, Long J, Zeng C, et al. Fine-scale mapping of the 4q24 locus identifies two independent loci associated with breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2015;24:1680–91.
13. Glubb DM, Maranian MJ, Michailidou K, 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;96:5–20.
14. Orr N, Dudbridge F, Dryden N, et al. Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. *Hum Mol Genet* 2015; 24:2966–84.
15. Dunning AM, Michailidou K, Kuchenbaecker KB, et al. Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nat Genet* 2016; 48:374–86.
16. Shi J, Zhang Y, Zheng W, et al. Fine-scale mapping of 8q24 locus identifies multiple independent risk variants for breast cancer. *Int J Cancer* 2016;139:1303–1317.
17. Horne HN, Chung CC, Zhang H, et al. Fine-mapping of the 1p11.2 breast cancer susceptibility locus. *PloS One* 2016;11:e0160316
18. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* 2007;81:208–27.
19. Rudolph A, Milne RL, Truong T, et al. Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors. *Int J Cancer* 2015;136:E685–96.
20. Nickels S, Truong T, Hein R, et al. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. *PLoS Genet* 2013;9: e1003284.
21. Barrdahl M, Canzian F, Joshi AD, et al. Post-GWAS gene-environment interplay in breast cancer: results from the Breast and Prostate Cancer Cohort Consortium and a meta-analysis on 79,000 women. *Hum Mol Genet* 2014;23:5260–70.
22. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst* 2011;103:1252–63.
23. Baer HJ, Rich-Edwards JW, Colditz GA, et al. Adult height, age at attained height, and incidence of breast cancer in premenopausal women. *Int J Cancer* 2006;119:2231–5.
24. Bassuk SS, Manson JE. Oral contraceptives and menopausal hormone therapy: relative and attributable risks of cardiovascular disease, cancer, and other health outcomes. *Ann Epidemiol* 2014;25: 193–200.
25. Beaber EF, Malone KE, Tang MT, et al. Oral contraceptives and breast cancer risk overall and by molecular subtype among young women. *Cancer. Epidemiol Biomarker Prev* 2014;23:755–64.
26. Beral V, Million Women Study C. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 2003;362:419–27.
27. Calle EE, Feigelson HS, Hildebrand JS, et al. Postmenopausal hormone use and breast cancer associations differ by hormone regimen and histologic subtype. *Cancer* 2009;115:936–45.
28. 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;13:1141–51.
29. Fagherazzi G, Vilier A, Boutron-Ruault MC, et al. Alcohol consumption and breast cancer risk subtypes in the E3N-EPIC cohort. *Eur J Cancer Prev* 2015;24:209–14.
30. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. *Epidemiol Rev* 1993;15: 36–47.
31. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell* 2013;155:934–47.
32. Sharma N, Thomas S, Golden EB, et al. Inhibition of autophagy and induction of breast cancer cell death by mefloquine, an antimalarial agent. *Cancer Lett* 2012;326:143–54.
33. Tanaka K, Whelan KA, Chandramouleeswaran PM, et al. ALDH2 modulates autophagy flux to regulate acetaldehyde-mediated toxicity thresholds. *Am J Cancer Res* 2016;6:781–96.
34. Sanchez-Alvarez R, Martinez-Outschoorn UE, Lin Z, et al. Ethanol exposure induces the cancer-associated fibroblast phenotype and lethal tumor metabolism: implications for breast cancer prevention. *Cell Cycle* 2013;12:289–301.
35. Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. *Cancer Epidemiol Biomarker Prev* 1999;8:1043–50.