Advancing research:
One cell at a time
One scientist at a time
One discovery at a time
Gene–environment interactions involving functional variants: Results from the Breast Cancer Association Consortium

Myrto Barndahl1, Anja Rudolph1, John L. Hopper2, Melissa C. Southey3, Annegien Broeks4, Peter A. Fasching5,6, Matthias W. Beckmann5, Manuela Gago-Dominguez7,8, J. Esteban Castelao8, Pascal Guéné6, Thérèse Truong10, Stig E. Bojesen11,12,13, Susan M. Gapstur14, Mia M. Gaudet15, Hermann Brenner15,16,17, Volker Arndt15, Hiltrud Brauch17,18,19, Ute Hamann20, Arto Mannemaa21,22,23, Diether Lambrechts24,25, Lynn Jongen26, Dieter Flesch-Jansys27,28, Kathrin Thoene28, Fergus J. Couch29, Graham G. Giles3,30, Jacques Simard31, Mark S. Goldberg32,33, Jonine Figueroa34,35, Kyriaki Michailidou36,37, Manjeet K. Bolla36, Joe Dennis36, Qin Wang36, Ursula Eibler1, Sabine Behrens1, Kamila Czene38, Per Hall38, Angela Cox39, Simon Cross40, Anthony Swerdlow41,42, Minouk J. Schoemaker42, Alison M. Dunning43, Rudolf Kaaks1, Paul D.P. Pharoah36,39, Marjanka Schmidt44,45,Montserrat García-Closas35, Douglas F. Easton36,43, Roger L. Milne3,30 and Jenny Chang-Claude1,45

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: CI1287/A16563, CI1287/A10118, CI1287/A10710, CI12292/A11174, CI1281/A12014, CS047/A8384, CS047/A15007, CS047/ A10692, CB197/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: U51 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 number: 1-117-1434-Hic; Grant sponsor: KAU; Grant number: 1-117-1434-Hic; 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: Conselleria de Sanidad: Conselleria de Economia Programa Sectorial de Investigación Aplicada, PEME 1 + D e 1 + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Conselleria 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 Economía 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, 01KW9976/6, 01KW9977/0, 01KW0014; 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: Grohe Family Foundation; Grant sponsor: VicHealth; Grant sponsor: Cancer Council Victoria; Grant sponsor: Cancer Genetics and Epigenetics

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 worthiness 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 (ORint = 0.77, 95% CI: 0.67–0.88, pint = 1.8 × 10\(^{-5}\)). The interaction with the strongest statistical evidence was found between 5q14-rs7707921 and alcohol consumption (ORint = 1.36, 95% CI: 1.16–1.59, pint = 1.9 × 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 (ORint = 1.26, 95% CI: 1.12–1.43, pint = 1.8 × 10\(^{-5}\)) and between 8q23-rs13267382 and age at first full-term pregnancy (ORint = 0.89, 95% CI: 0.83–0.95, pint = 5.2 × 10\(^{-5}\)). 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.
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. 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. 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 ≥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/m²), 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 progesterone) 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 × 10$^{-5}$). Interaction ORs were tested for heterogeneity across studies based on 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. 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 × 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$_{per-allele}$ = 0.76, 95% CI: 0.66–
0.88, \( p = 2.2 \times 10^{-4} \) compared to that of non-smoker carriers (\( OR_{\text{per-allele}} = 0.99, 95\% \text{ CI: } 0.91-1.08, p = 0.9 \)) where no risk association was observed. When comparing ER– cases to ER+ cases, the results did not indicate any effect heterogeneity (\( p_{\text{het}} = 0.48 \)). There was no strong evidence of interaction, neither with respect to ER+ risk (\( p_{\text{int}} = 0.0014 \)) nor with respect to ER– risk (\( p_{\text{int}} = 0.75 \)).

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 (\( OR_{\text{int}} = 1.36, 95\% \text{ CI: } 1.16-1.59, p_{\text{int}} = 1.9 \times 10^{-5} \)) 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 \(<20 \text{ g of alcohol per day (OR}_{\text{per-allele}} = 1.07, 95\% \text{ CI: } 0.92-1.24, p = 0.36 \). A strong effect heterogeneity was detected when comparing ER– cases to ER+ cases (\( p_{\text{het}} = 6.7 \times 10^{-5} \)). Together with the absence of interaction in relation to ER+ disease (\( p_{\text{int}} = 0.79 \)) and overall breast cancer risk (\( p_{\text{int}} = 0.70 \)), 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 (\( OR_{\text{int}} = 1.26, 95\% \text{ CI: } 1.12-1.43, p_{\text{int}} = 1.8 \times 10^{-4} \)) 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 (\( OR_{\text{per-allele}} = 0.70, 95\% \text{ CI: } 0.54-0.90, p = 0.006 \), 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 (\( OR_{\text{per-allele}} = 0.88,\) \( p = 0.20 \)).
95% CI: 0.76–1.02, \( p = 0.08 \), or after the age of 14 years (OR_{per-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 (\( p_{\text{het}} = 0.53 \)) nor was there any indication of any interaction in relation to overall breast cancer risk (\( p_{\text{int}} = 0.94 \)). 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, \( p_{\text{int}} = 5.2 \times 10^{-5} \)) 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 (\( p_{\text{int}} = 0.98 \)), or overall breast cancer risk (\( p_{\text{int}} = 0.47 \)), and no effect heterogeneity was found when comparing the risk of ER– and ER+ breast cancer (\( p_{\text{het}} = 0.99 \)). 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-allele} = 0.79, 95% CI: 0.68–0.91, \( p = 0.001 \)), whereas for women who had their first child at a younger age the allele had no effect on risk.

**Discussion**

From the analyses presented in this work, four SNP-risk factor pairs were identified, for which \( p_{\text{int}} < 7 \times 10^{-4} \), 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 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 (\( p_{\text{int}} = 1.8 \times 10^{-5} \)). 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.\(^5\)\(^3\)\(^1\) 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 (\( p_{\text{int}} = 1.9 \times 10^{-5} \)). 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.\(^3\)\(^2\) It has been reported that autophagy could have a protective effect on esophageal epithelial cells responding to ethanol-induced oxidative stress.\(^3\)\(^3\) Also, while ethanol promotes oxidative stress in cancer associated fibroblasts, it has been reported to induce autophagy resistance in epithelial cells.\(^3\)\(^4\) 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 (\( p_{\text{int}} = 1.8 \times 10^{-4} \)) 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 \( \geq 14 \) 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 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 (\( p_{\text{int}} = 2.6 \times 10^{-7} \)) 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>
<thead>
<tr>
<th>SNP/risk factor</th>
<th>Stratum</th>
<th>Cases/controls</th>
<th>OR (95% CI)</th>
<th>( P_{\text{interaction}} )</th>
<th>OR (95% CI)</th>
<th>( P_{\text{interaction}} )</th>
<th>OR (95% CI)</th>
<th>( P_{\text{interaction}} )</th>
<th>( P_{\text{case-only}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFLAR-rs7558475/current smoking</td>
<td>No</td>
<td>7,698/8,835</td>
<td>0.99 (0.91–1.08)</td>
<td>1.00 (0.91–1.09)</td>
<td>0.96 (0.81–1.14)</td>
<td>1.8 ( 10^{-10} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1,375/1,519</td>
<td>0.76 (0.66–0.88)</td>
<td>0.78 (0.66–0.91)</td>
<td>0.91 (0.63–1.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>9,073/10,354</td>
<td>0.96 (0.88–1.04)</td>
<td>1.8 ( 10^{-10} )</td>
<td>0.96 (0.82–1.12)</td>
<td>0.75</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5q14-rs7707921/alcohol consumption</td>
<td>(&lt;20 \text{ g/day})</td>
<td>4,904/5,411</td>
<td>1.16 (1.08–1.24)</td>
<td>1.18 (1.10–1.27)</td>
<td>1.07 (0.92–1.24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\geq20 \text{ g/day})</td>
<td>481/541</td>
<td>1.16 (0.94–1.43)</td>
<td>1.08 (0.86–1.36)</td>
<td>2.59 (1.45–4.62)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>5,385/5,952</td>
<td>1.16 (1.09–1.23)</td>
<td>0.70</td>
<td>1.15 (0.99–1.32)</td>
<td>1.9 ( 10^{-5} )</td>
<td>6.7 ( 10^{-6} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3p21-rs6796502/age at menarche</td>
<td>(\leq11 \text{ years})</td>
<td>3,350/3,609</td>
<td>0.73 (0.65–0.83)</td>
<td>0.75 (0.65–0.86)</td>
<td>0.70 (0.54–0.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12–13 years</td>
<td>9,503/10,893</td>
<td>0.93 (0.86–0.99)</td>
<td>0.93 (0.86–1.00)</td>
<td>0.88 (0.76–1.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\geq14 \text{ years})</td>
<td>7,294/8,864</td>
<td>0.95 (0.88–1.03)</td>
<td>0.92 (0.84–1.01)</td>
<td>1.16 (0.99–1.34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>20,147/23,366</td>
<td>0.90 (0.86–0.95)</td>
<td>0.94</td>
<td>0.94 (0.86–1.04)</td>
<td>1.8 ( 10^{-4} )</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8q23-rs13267382/age at first FTP</td>
<td>(&lt;20 \text{ years})</td>
<td>2,085/1,830</td>
<td>1.01 (0.92–1.11)</td>
<td>1.00 (0.90–1.11)</td>
<td>1.12 (0.94–1.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20–24 years</td>
<td>6,944/8,246</td>
<td>0.92 (0.88–0.97)</td>
<td>0.91 (0.86–0.96)</td>
<td>1.00 (0.91–1.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25–29 years</td>
<td>5,388/6,877</td>
<td>0.97 (0.92–1.02)</td>
<td>0.98 (0.92–1.03)</td>
<td>0.91 (0.81–1.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\geq30 \text{ years})</td>
<td>2,965/3,555</td>
<td>0.92 (0.85–0.99)</td>
<td>0.94 (0.87–1.02)</td>
<td>0.79 (0.68–0.91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>17,382/20,508</td>
<td>0.94 (0.92–0.97)</td>
<td>0.47</td>
<td>0.95 (0.89–1.01)</td>
<td>5.2 ( 10^{-4} )</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 (20,147 cases and 23,366 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. Eales, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRAC-TICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Andrew Lee and Ed Dicks, Craig Laccarini 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 Genome 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. Hlikker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. Australian Breast Cancer Family Study thanks Maggie Angelakos, Judi Makiell 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, Emil Rutgers, C Ellen van der Schoot and Femke Atsma. Breast Oncology Genetics 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 Xestión 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: Dorte Uldall Andersen, Maria Birna Arnadottir, Anne Bank and Dorthe Kjeldgaard 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 Herrmann, Christa Stegmayer 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 Kermiläinen. Leuven Multidisciplinary Breast Centre thanks Gilian Peuteeman, Dominiek Smeets, Thomas
Van Brussel and Kathleen Courthouts. Mammary Carcinoma Risk Factor Investigation thanks Petra Seebold, Judith Heinz, Nadia Obi, Alina Vrielings, Muhabbet Celik, Til Ochlers and Stefan Nickels. Montreal Gene-Environment Breast Cancer Study thanks Martine Tranchant (CHU de Québec Research Center), Marie-France Valsos, 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, Neolina Szczesnia-Dabrowska, Beata Peplonska, Witold Zatonksi, 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 Comnley. 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 this work. This study was supported by NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre as well as multiple funding agencies: funding for BCAC and the iCOGS infrastructure came from: Cancer Research UK (C1287/A16563, C1287/A10118, C1287/A10710, C12292/A11714, C1281/A12014, CS047/A8384, CS047/A15007, CS047/A10692, CB197/A16655), 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 (HEALTHFP2-2009-223175) (COCGs), 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 U1M1 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 the European Community’s Seventh Framework Programme under grant agreement no. 223175 (HEALTHFP2-2009-223175) (COCGs), 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 U1M1 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 the European Community’s Seventh Framework Programme under grant agreement no. 223175 (HEALTHFP2-2009-223175) (COCGs).
Cancer Genetics and Epigenetics

References