



This is a repository copy of *A genome-wide gene-environment interaction study of breast cancer risk for women of European ancestry*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/202417/>

Version: Published Version

---

**Article:**

Middha, P., Wang, X., Behrens, S. et al. (129 more authors) (2023) A genome-wide gene-environment interaction study of breast cancer risk for women of European ancestry. *Breast Cancer Research*, 25. 93.

<https://doi.org/10.1186/s13058-023-01691-8>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

RESEARCH

Open Access



# A genome-wide gene-environment interaction study of breast cancer risk for women of European ancestry

Pooja Middha<sup>1\*</sup>, Xiaoliang Wang<sup>2,3</sup>, Sabine Behrens<sup>1</sup>, Manjeet K. Bolla<sup>4</sup>, Qin Wang<sup>4</sup>, Joe Dennis<sup>4</sup>, Kyriaki Michailidou<sup>4,5</sup>, Thomas U. Ahearn<sup>6</sup>, Irene L. Andrulis<sup>7,8</sup>, Hoda Anton-Culver<sup>9</sup>, Volker Arndt<sup>10</sup>, Kristan J. Aronson<sup>11</sup>, Paul L. Auer<sup>12</sup>, Annelie Augustinsson<sup>13</sup>, Thäis Baert<sup>14</sup>, Laura E. Beane Freeman<sup>6</sup>, Heiko Becher<sup>15</sup>, Matthias W. Beckmann<sup>16</sup>, Javier Benitez<sup>17,18</sup>, Stig E. Bojesen<sup>19,20,21</sup>, Hiltrud Brauch<sup>22,23,24</sup>, Hermann Brenner<sup>10,25,26</sup>, Angela Brooks-Wilson<sup>27</sup>, Daniele Campa<sup>1,28</sup>, Federico Canzian<sup>29</sup>, Angel Carracedo<sup>30,31</sup>, Jose E. Castelao<sup>32</sup>, Stephen J. Chanock<sup>6</sup>, Georgia Chenevix-Trench<sup>33</sup>, CTS Consortium<sup>34,35</sup>, Emilie Cordina-Duverger<sup>36</sup>, Fergus J. Couch<sup>37</sup>, Angela Cox<sup>38</sup>, Simon S. Cross<sup>39</sup>, Kamila Czene<sup>40</sup>, Laure Dossus<sup>41</sup>, Pierre-Antoine Dugué<sup>42,43</sup>, A. Heather Eliassen<sup>44,45,46</sup>, Mikael Eriksson<sup>40</sup>, D. Gareth Evans<sup>47,48</sup>, Peter A. Fasching<sup>16</sup>, Jonine D. Figueroa<sup>6,49,50</sup>, Olivia Fletcher<sup>51</sup>, Henrik Flyger<sup>52</sup>, Marike Gabrielson<sup>40</sup>, Manuela Gago-Dominguez<sup>30</sup>, Graham G. Giles<sup>42,43,53</sup>, Anna González-Neira<sup>54</sup>, Felix Grassmann<sup>40,55</sup>, Anne Grundy<sup>56</sup>, Pascal Guéné<sup>36</sup>, Christopher A. Haiman<sup>57</sup>, Niclas Håkansson<sup>58</sup>, Per Hall<sup>40,59</sup>, Ute Hamann<sup>60</sup>, Susan E. Hankinson<sup>44,61</sup>, Elaine F. Harkness<sup>62,63,64</sup>, Bernd Hollecsek<sup>65</sup>, Reiner Hoppe<sup>22,66</sup>, John L. Hopper<sup>53</sup>, Richard S. Houlston<sup>67</sup>, Anthony Howell<sup>68</sup>, David J. Hunter<sup>45,69</sup>, Christian Ingvar<sup>70</sup>, ABCTB Investigators<sup>71</sup>, kConFab Investigators<sup>72,73</sup>, Karolin Isaksson<sup>74</sup>, Helena Jernström<sup>13</sup>, Esther M. John<sup>75,76</sup>, Michael E. Jones<sup>67</sup>, Rudolf Kaaks<sup>1</sup>, Renske Keeman<sup>77</sup>, Cari M. Kitahara<sup>78</sup>, Yon-Dschun Ko<sup>79</sup>, Stella Koutros<sup>6</sup>, Allison W. Kurian<sup>75,76</sup>, James V. Lacey<sup>34,35</sup>, Diether Lambrechts<sup>80,81</sup>, Nicole L. Larson<sup>82</sup>, Susanna Larsson<sup>58,83</sup>, Loic Le Marchand<sup>84</sup>, Flavio Lejbkowicz<sup>85</sup>, Shuai Li<sup>4,42,53</sup>, Martha Linet<sup>78</sup>, Jolanta Lissowska<sup>86</sup>, Maria Elena Martinez<sup>87</sup>, Tabea Maurer<sup>88</sup>, Anna Marie Mulligan<sup>89,90</sup>, Claire Mulot<sup>91</sup>, Rachel A. Murphy<sup>92,93</sup>, William G. Newman<sup>47,48</sup>, Sune F. Nielsen<sup>19,20</sup>, Børge G. Nordestgaard<sup>19,20,21</sup>, Aaron Norman<sup>82</sup>, Katie M. O'Brien<sup>94</sup>, Janet E. Olson<sup>82</sup>, Alpa V. Patel<sup>95</sup>, Ross Prentice<sup>96</sup>, Erika Rees-Punia<sup>95</sup>, Gad Rennert<sup>85</sup>, Valerie Rhenius<sup>97</sup>, Kathryn J. Ruddy<sup>98</sup>, Dale P. Sandler<sup>94</sup>, Christopher G. Scott<sup>99</sup>, Mitul Shah<sup>97</sup>, Xiao-Ou Shu<sup>100</sup>, Ann Smeets<sup>101</sup>, Melissa C. Southey<sup>42,43,102</sup>, Jennifer Stone<sup>53,103</sup>, Rulla M. Tamimi<sup>45,104</sup>, Jack A. Taylor<sup>94,105</sup>, Lauren R. Teras<sup>95</sup>, Katarzyna Tomczyk<sup>51</sup>, Melissa A. Troester<sup>106</sup>, Thérèse Truong<sup>36</sup>, Celine M. Vachon<sup>82</sup>, Sophia S. Wang<sup>34,35</sup>, Clarice R. Weinberg<sup>107</sup>, Hans Wildiers<sup>14</sup>, Walter Willett<sup>44,45,46</sup>, Stacey J. Winham<sup>108</sup>, Alicja Wolk<sup>58,83</sup>, Xiaohong R. Yang<sup>6</sup>, M. Pilar Zamora<sup>109</sup>, Wei Zheng<sup>100</sup>, Argyrios Ziogas<sup>9</sup>, Alison M. Dunning<sup>97</sup>, Paul D. P. Pharoah<sup>4,97</sup>, Montserrat García-Closas<sup>6</sup>, Marjanka K. Schmidt<sup>77,110</sup>, Peter Kraft<sup>45,111</sup>, Roger L. Milne<sup>42,43,53</sup>, Sara Lindström<sup>2,3</sup>, Douglas F. Easton<sup>4,97†</sup> and Jenny Chang-Claude<sup>1,88†</sup>

<sup>†</sup>Douglas F. Easton, Jenny Chang-Claude Joint senior authors.

\*Correspondence:

Pooja Middha

pooja.middha@ucsf.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Abstract

**Background** Genome-wide studies of gene–environment interactions (G×E) may identify variants associated with disease risk in conjunction with lifestyle/environmental exposures. We conducted a genome-wide G×E analysis of ~7.6 million common variants and seven lifestyle/environmental risk factors for breast cancer risk overall and for estrogen receptor positive (ER+) breast cancer.

**Methods** Analyses were conducted using 72,285 breast cancer cases and 80,354 controls of European ancestry from the Breast Cancer Association Consortium. Gene–environment interactions were evaluated using standard unconditional logistic regression models and likelihood ratio tests for breast cancer risk overall and for ER+ breast cancer. Bayesian False Discovery Probability was employed to assess the noteworthiness of each SNP–risk factor pairs.

**Results** Assuming a  $1 \times 10^{-5}$  prior probability of a true association for each SNP–risk factor pairs and a Bayesian False Discovery Probability < 15%, we identified two independent SNP–risk factor pairs: rs80018847(9p13)-*LINGO2* and adult height in association with overall breast cancer risk ( $OR_{int} = 0.94$ , 95% CI 0.92–0.96), and rs4770552(13q12)-*SPATA13* and age at menarche for ER+ breast cancer risk ( $OR_{int} = 0.91$ , 95% CI 0.88–0.94).

**Conclusions** Overall, the contribution of G×E interactions to the heritability of breast cancer is very small. At the population level, multiplicative G×E interactions do not make an important contribution to risk prediction in breast cancer.

**Keywords** Breast cancer, Gene–environment interactions, Genetic epidemiology, European ancestry

## Background

Breast cancer is a complex disease involving interplay between lifestyle/environmental and genetic risk factors. Risk factors such as parity, breastfeeding, age at menarche, age at first full-term pregnancy, body mass index (BMI), height, mammographic density, exogenous hormonal use, and alcohol consumption are well-established [1–7]. Through continued collaborative efforts such as the Collaborative Oncological Gene–environment Study (COGS) and the OncoArray project [8], more than 200 common single nucleotide polymorphisms (SNPs) associated with risk of breast cancer have been identified [9–11].

Traditional genome-wide association study (GWAS) analyses assess the marginal effects of variants and might miss variants which only show an effect within certain strata in the population. These potential gene–environment interactions where SNPs are associated with disease risk in conjunction with lifestyle/environmental risk factors can be investigated through genome-wide gene–environment interaction studies (GEWIS) [12–15].

Very few genome-wide studies of gene–environment (G×E) interactions in breast cancer have been conducted to date, and three focused on the use of menopausal hormonal therapy as the single environmental risk factor [16–18]. An exploratory analysis of G×E interactions examined ten environmental risk factors and 71,527 SNPs selected from prior evidence, using data from approximately 35,000 cases and controls in the Breast Cancer Association Consortium (BCAC). That study identified two potential G×E interactions

associated with breast cancer risk [19]. In the present study, we performed a comprehensive genome-wide analysis of gene–environment interactions for risk of overall breast cancer, as well as estrogen receptor positive (ER+) breast cancer using data from 72,285 cases and 80,354 controls participating in the BCAC.

## Methods

### Study sample

Analyses were conducted using data from 46 studies (16 prospective cohorts, 14 population-based case–control studies, and 16 non-population based studies) participating in the BCAC. We excluded participants if they were genotypically male, of non-European descent, or had a breast tumor of unknown invasiveness or in-situ breast cancer. Women with prevalent breast cancer at the time of recruitment or with unknown reference age (defined as age at diagnosis for cases and age at interview for controls) were also excluded from the analyses. Further, studies with fewer than 150 cases and 150 controls for the risk factor under evaluation were excluded from those analyses. Each participating study obtained informed consent from the participants and was approved by their local ethics committee.

### Risk factor data

Risk factor data from individual studies was checked for quality using a multi-step harmonization process based on a common data dictionary. Time-dependent risk factor variables were derived with respect to the reference date defined as date of diagnosis for cases and date of interview for controls. Analyses were conducted with the

following risk factors among all women: age at menarche (per 2 years), parity (per 1 birth), adult height (per 5 cm), ever use of oral contraceptives (yes/no), and current smoking (yes/no). The analysis of age at first full-term pregnancy (per 5 years) was conducted among parous women only, and that of body mass index (BMI, per 5 kg/m<sup>2</sup>) was conducted among postmenopausal women only. Menopausal status was either self-reported or assigned as postmenopausal if the reference age was greater than 54 years.

### Genetic data

All samples were genotyped either using the iCOGS [20, 21] or OncoArray [9, 10, 22]. Briefly, iCOGS is a customized iSelect SNP genotyping array, consisting of ~211,000 SNPs [20, 21], whereas OncoArray includes ~533,000 SNPs of which nearly 260,000 were selected as a GWAS backbone (Illumina HumanCore) [22]. Detailed information is provided elsewhere [9, 10, 20–22]. Data were imputed to the 1000 Genomes Reference Panel (phase 3 version 5). Overall, 28,176 cases and 32,209 controls of European ancestry who were genotyped by the iCOGS array, and 44,109 cases and 48,145 controls who were genotyped using the OncoArray array were included in this analysis.

Genetic variants with imputation quality score <0.5 in iCOGS or <0.8 in OncoArray, or with minor allele frequency <0.01, were excluded from the analyses. Variants in known breast cancer regions were also excluded from the analysis since interactions between known susceptibility variants and risk factors have been explored previously [23, 24]. After applying all exclusions, 7,672,870 genetic variants (SNPs and indels) were included in the analysis.

### Statistical analysis

Unconditional logistic regression was employed to assess the associations of SNPs and risk factors with breast cancer risk. Genotypes were assessed using the expected number of copies of the alternative allele ('dosage') as the covariate under a log-additive model. Interactions between genetic variants and risk factors were tested by comparing the fit of logistic regression models with and without an interaction term using likelihood ratio tests. All models were adjusted for reference age, study, and ten ancestry-informative principal components. To account for potential differential main effects of risk factors by study design, all models included an interaction term between risk factor and an indicator variable for study design (population-based vs. non-population based). Analyses with current smoking were further adjusted for former smoking.

Analyses were performed separately for overall and ER+ breast cancer risk, and also separately by genotyping array. Array-specific results were combined using METAL [25]. Quantile–quantile (Q-Q) plots were assessed to examine the consistency of the distribution of p-values with the null distribution. Interaction P value less than 5E-07 was considered suggestive evidence of interaction. We also calculated Bayesian False Discovery Probabilities (BFDP) for all suggestive interactions, assuming a  $1 \times 10^{-5}$  prior probability of a true association for each SNP-risk factor pair. Overall, G×E interactions with BFDP <15% were considered noteworthy [26]. For noteworthy SNP-risk pairs, we evaluated the G×E interaction also for ER-negative breast cancer risk. For noteworthy interactions, we conducted stratified analyses by categories of the risk factor. All analyses were conducted using R version 3.5.1.

We estimated the overall genome-wide contribution of G×E associations for each risk factor to the familial relative risk of breast cancer using LD score regression [27]. The analysis used the G×E interaction summary statistics and was restricted to HapMap3 SNPs with MAF >5% in European population from the 1000 Genomes Project. Under the log-additive model, the G×E heritability on the frailty scale can be estimated by  $h_f^2 = h_{obs}^2 \times var(X)/P(1-P)$ , where  $h_{obs}^2$  is the observed heritability given by LD score regression,  $var(X)$  is the variance of the risk factor under evaluation, and  $P$  is the proportion of cases in the sample. The proportion of the familial relative risk (FRR) of breast cancer due to G×E interactions is then given by  $h_f^2/2\log(\lambda)$  where  $\lambda$  is the familial relative risk to first degree relatives of cases (assumed to be 2) [28].

### Results

Studies included in the analysis are summarized in Additional file 1: Table S1. The number of cases and controls in each analysis varied from 61,617 cases and 74,698 controls for parity to 48,276 cases and 60,587 controls for current smoking (Additional file 1: Table S2). Consistent with the literature, increasing age at first full-term pregnancy, higher adult height, ever use of oral contraceptives, and current smoking were associated with increased overall breast cancer risk, whereas increasing age at menarche, being parous, increasing number of full-term pregnancies, and breast feeding were associated with decreased breast cancer risk (Additional file 1: Table S3).

The genome-wide analysis of interactions with seven environmental risk factors yielded two SNP-risk factor pairs at BFDP <15%, one for risk of overall breast cancer and one for ER+ breast cancer risk (Table 1, Fig. 1,

**Table 1** Genetic variants with suggestive ( $P_{\text{int}} \leq 5E-07$ ) GxE interactions for overall and estrogen receptor positive (ER+) breast cancer risk

Risk factor	SNP	Chr	Position <sup>1</sup>	Alleles <sup>2</sup>	EAF <sup>3</sup>	OR <sub>marg</sub> (95% CI)	P <sub>marg</sub>	OR <sub>int</sub> (95% CI)	P <sub>int</sub>	BFD <sup>4</sup>
<i>Overall breast cancer risk</i>										
Number of full-term pregnancies (per 1 birth)	rs10928872	2	129,833,111	A/T	0.08	1.00 (0.97–1.03)	0.97	0.93 (0.91–0.96)	1.34E-07	0.98
Number of full-term pregnancies (per 1 birth)	rs36064687	2	129,832,988	AT/A	0.08	1.00 (0.97–1.04)	0.99	0.93 (0.91–0.96)	1.34E-07	0.98
Number of full-term pregnancies (per 1 birth)	rs79929694	2	129,841,483	A/G	0.08	1.00 (0.96–1.03)	0.96	0.93 (0.91–0.96)	1.43E-07	0.98
Number of full-term pregnancies (per 1 birth)	rs77107485	2	129,843,663	T/G	0.08	1.00 (0.97–1.04)	0.89	0.93 (0.91–0.96)	2.79E-07	0.98
Number of full-term pregnancies (per 1 birth)	rs79722231	2	129,834,483	T/C	0.08	1.00 (0.97–1.04)	0.92	0.93 (0.91–0.96)	2.80E-07	0.98
Age at menarche (per 2 years)	rs73277506	7	21,112,413	C/T	0.02	1.03 (0.96–1.10)	0.39	1.26 (1.16–1.38)	1.46E-07	0.75
Current smoking (yes/no)	rs11322161	8	105,120,195	GC/G	0.29	1.00 (0.98–1.02)	0.63	1.19 (1.12–1.27)	9.26E-08	0.49
Adult height (per 5 cm)	rs80018847	9	28,326,896	A/G	0.14	1.00 (0.98–1.03)	0.88	0.94 (0.92–0.96)	4.34E-08	0.11
Adult height (per 5 cm)	rs1360506	9	28,339,154	G/C	0.16	1.00 (0.97–1.03)	0.96	0.95 (0.93–0.98)	2.13E-07	1.00
Adult height (per 5 cm)	rs1237669	17	41,912,024	T/G	0.91	1.01 (0.98–1.04)	0.62	0.94 (0.91–0.96)	5.00E-07	0.16
OC use (yes/no)	rs147290549	18	7,713,860	C/T	0.50	1.01 (0.99–1.02)	0.23	1.10 (1.06–1.14)	4.55E-07	0.59
OC use (yes/no)	rs664040	18	7,716,250	C/A	0.50	1.01 (0.99–1.02)	0.22	1.10 (1.06–1.14)	4.93E-07	0.59
Current smoking (yes/no)	rs75489324	21	26,302,665	G/C	0.02	0.98 (0.92–1.05)	0.87	0.58 (0.47–0.72)	4.58E-07	0.94
<i>ER+ breast cancer risk</i>										
Age at menarche (per 2 years)	rs73277506	7	21,112,413	C/T	0.02	1.04 (0.97–1.12)	0.25	1.28 (1.17–1.41)	3.94E-07	0.76
Current smoking (yes/no)	rs11322161	8	105,120,195	GC/G	0.29	1.00 (0.98–1.02)	0.95	1.20 (1.12–1.29)	4.32E-07	0.81
Age at menarche (per 2 years)	rs4770552	13	24,594,430	C/T	0.86	1.02 (1.00–1.05)	0.10	0.91 (0.88–0.94)	4.62E-08	0.11
Age at menarche (per 2 years)	rs113684695	13	24,600,947	ACCTCGT GATCCGC/A	0.86	1.02 (1.00–1.05)	0.11	0.91 (0.88–0.94)	5.48E-08	0.11
Age at menarche (per 2 years)	rs9510997	13	24,592,857	G/T	0.86	1.02 (1.00–1.05)	0.10	0.91 (0.88–0.94)	6.26E-08	0.11

**Table 1** (continued)

Risk factor	SNP	Chr	Position <sup>1</sup>	Alleles <sup>2</sup>	EAF <sup>3</sup>	OR <sub>marg</sub> (95% CI)	P <sub>marg</sub>	OR <sub>int</sub> (95% CI)	P <sub>int</sub>	BFDP <sup>4</sup>
Age at menarche (per 2 years)	rs7321200	13	24,593,519	A/G	0.87	1.01 (0.99–1.04)	0.33	0.91 (0.87–0.94)	7.92E-08	0.11
Age at menarche (per 2 years)	rs4770553	13	24,595,995	C/T	0.85	1.02 (0.99–1.04)	0.22	0.92 (0.89–0.95)	4.31E-07	0.76
Age at menarche (per 2 years)	rs9551041	13	24,601,810	A/G	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.58E-07	0.76
Age at menarche (per 2 years)	rs1886805	13	24,602,494	G/A	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.62E-07	0.76
Age at menarche (per 2 years)	rs4770555	13	24,601,751	T/C	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.63E-07	0.76
Age at menarche (per 2 years)	rs4770556	13	24,601,760	A/G	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.68E-07	0.76
Age at menarche (per 2 years)	rs1886804	13	24,602,174	T/G	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.72E-07	0.76
Age at menarche (per 2 years)	rs1886803	13	24,602,165	T/C	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.72E-07	0.76
Age at menarche (per 2 years)	rs9553145	13	24,602,061	G/C	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.75E-07	0.76
Age at menarche (per 2 years)	rs4769305	13	24,592,448	T/C	0.85	1.02 (0.99–1.05)	0.17	0.92 (0.89–0.95)	4.76E-07	0.76
Age at menarche (per 2 years)	rs9553144	13	24,601,943	C/G	0.85	1.02 (0.99–1.04)	0.21	0.92 (0.89–0.95)	4.78E-07	0.76
Number of full-term pregnancies (per 1 birth)	rs930198	23	79,492,077	T/C	0.24	1.00 (0.98–1.02)	0.93	0.96 (0.94–0.97)	3.10E-07	0.99

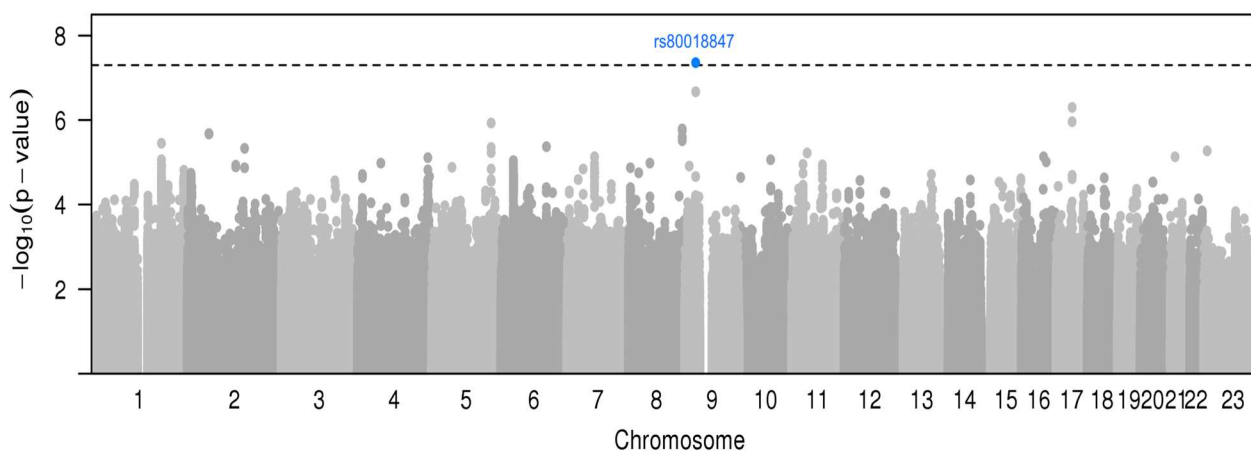
SNP: Single Nucleotide Polymorphism; Chr: Chromosome; Ref: Reference Allele; Alt: Alternative Allele; EAF: Effect Allele Frequency; OR<sub>marg</sub>: SNP Marginal Odds ratio; CI: Confidence Interval; P<sub>marg</sub>: Marginal p-value (meta-analyzed); OR<sub>int</sub>: Interaction p-value (meta-analyzed); P<sub>int</sub>: Interaction p-value; BFDP: Bayesian False Discovery Probability; ABF: Approximate Bayes Probability; OC use: Ever use of oral contraceptives; ER+: Estrogen receptor positive breast cancer risk

<sup>1</sup> Build 37 Position

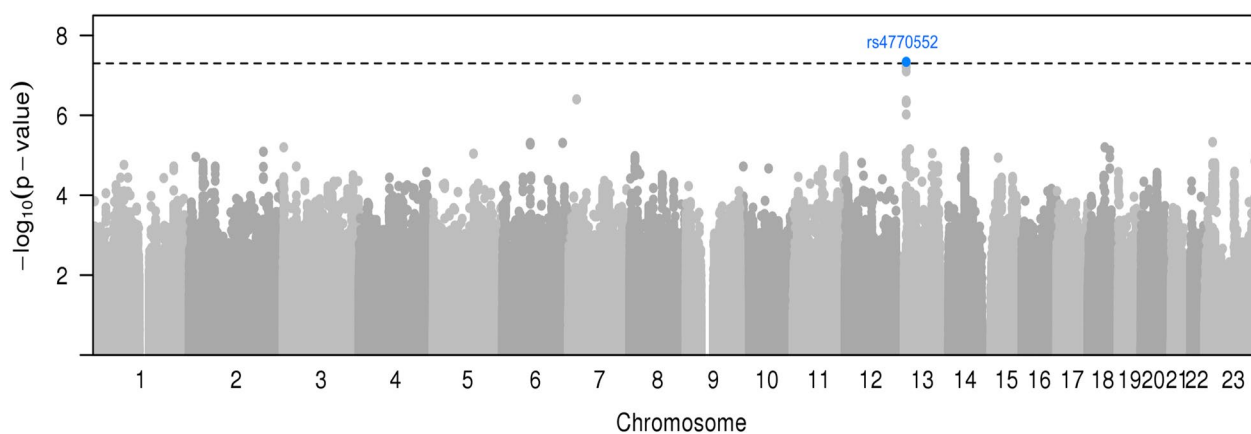
<sup>2</sup> Reference/Alternate alleles in Europeans (forward strand)

<sup>3</sup> Effect allele frequency in controls in OncoArray dataset

<sup>4</sup> Bayesian False Discovery Probability at prior probability of  $1 \times 10^{-05}$



**Fig. 1** Manhattan plot of genome-wide interactions of adult height on overall breast cancer risk. The genome-wide significance threshold of  $P < 5 \times 10^{-8}$  is indicated by the dashed black line. Genome-wide significant findings are highlighted in blue



**Fig. 2** Manhattan plot of genome-wide interaction of age at menarche for ER+ breast cancer risk. The genome-wide significance threshold of  $P < 5 \times 10^{-8}$  is indicated by the dashed black line. Genome-wide significant findings are highlighted in blue.

2, Additional file 1: Figure S1A-S1B). No inflation in the test statistics was observed for either of the environmental risk factors. The heritability on the frailty scale of breast cancer risk explained by G×E interaction is shown in Additional file 1: Figure S2. The estimated proportion of the frailty scale heritability explained by G×E interactions was very low for all factors, being highest for age at first full-term pregnancy (~1.5% for both overall and ER+ breast cancer risk), age at menarche and post-menopausal BMI.

For overall breast cancer risk, there was evidence of interaction between SNP rs80018847 and adult height ( $OR_{int}=0.94$ , 95% CI 0.92–0.96,  $P_{int}=4.34E-08$ , BFD $P=11\%$ ) without an apparent marginal effect of the rs80018847 variant ( $OR_{marg}=1.00$ , 95% CI 0.98–1.03,  $P_{marg}=0.88$ ). By categories of adult height defined a priori, the estimated per allele  $OR_{meta}$  of rs80018847-G varied from 1.03 (95% CI 0.94–1.13,  $P_{meta}=0.53$ ) for women

shorter than 158 cm, 1.13 (1.02–1.25) for women 158–162 cm in height, to  $OR_{meta}$  of 1.01 (95% CI 0.93–1.09,  $P_{meta}=0.88$ ) for women who were 168 cm or taller risk (Additional file 1: Table S4). Therefore, there is no linear relationship between the SNP and categories of adult height. The interaction with height was also observed for ER+ breast cancer ( $OR_{int}=0.95$ , 95% CI 0.93–0.97,  $P_{int}=5.62E-06$ ) but not for ER negative (ER-) breast cancer risk ( $OR_{int}=0.98$ , 95% CI 0.93–1.03,  $P_{int}=0.77$ ). The regional plot for overall breast cancer shows another SNP (rs1360506) at this locus in high linkage disequilibrium (LD) ( $r^2=0.81$ ) with rs80018847 (Additional file 1: Figure S3).

For risk of ER+ breast cancer, a statistically significant interaction was observed between SNP rs4770552 and age at menarche ( $OR_{int}=0.91$ , 95% CI 0.88–0.94,  $P_{int}=4.62E-08$ , BFD $P=11\%$ ). There was weak evidence for a marginal association between the rs4770552-T

allele and ER+breast cancer ( $OR_{\text{marg}}=1.02$ , 95% CI 1.00–1.05,  $P_{\text{marg}}=0.10$ ). The per allele  $OR_{\text{meta}}$  appeared to decrease with increasing age at menarche, from 1.07 (95% CI 1.00–1.15,  $P_{\text{meta}}=0.04$ ) for age at menarche less than 13 years to 0.92 (95% CI 0.77–1.09,  $P_{\text{meta}}=0.33$ ) for age at menarche greater than 15 years (Additional file 1: Table S4). There was weaker evidence of interaction between SNP rs4770552 and age at menarche for overall breast cancer risk ( $OR_{\text{int}}=0.93$ , 95% CI 0.90–0.96,  $P_{\text{int}}=5.47E-06$ ), but no interaction for ER- breast cancer risk ( $OR_{\text{int}}=0.98$ , 95% CI 0.89–1.08),  $P_{\text{int}}=0.66$ ). At this locus, we found suggestive evidence of interactions between further 13 SNPs and age at menarche for ER+breast cancer risk. However, these 13 SNPs are in high LD ( $r^2=0.8-1.0$ ) with SNP rs4770552 (Additional file 1: Figure S4).

## Discussion

This is the largest genome-wide gene-environment interaction study for breast cancer to date. We found evidence of one novel susceptibility loci interacting with adult height associated with increased breast cancer risk overall, and one interaction for increased risk of ER+breast cancer with age at menarche. It is important to note, however, that while these associations reached conventional levels of genome-wide statistical significance, they may still represent chance associations. Based on the assumed prior distribution of effect sizes, the BFD for both loci were 11%, considered noteworthy. Nevertheless, studies with an even larger sample size are required to confirm or refute these associations.

Many observational studies have shown an association between increasing adult height and increased breast cancer risk, in both premenopausal and postmenopausal women [7, 29, 30]. A meta-analysis estimated that each 10 cm increment in height was associated with a 17% increase in breast cancer risk [31]. The biological link between height and breast cancer is poorly understood, but some studies have suggested that increased height corresponds to more stem cells at risk of acquiring driver mutations [32]. Another hypothesis is that adult height could be a surrogate for nutritional intake, potentially implying a role for insulin-like growth factor 1 (*IGF1*) [33]. The functional basis of the potential interaction between adult height and the SNP rs80018847 is unclear. This SNP is in an intronic region of the leucine rich repeat and Ig domain containing 2 gene (*LINGO2*) on the short arm of chromosome 9 (9p13). This gene encodes a transmembrane protein belonging to the LINGO/LERN protein family [34]. Studies in mouse embryos have shown expression of *LINGO2* specifically in the central

nervous system [34], but it has not been implicated in breast cancer to date.

Early age at menarche is known to be associated with elevated risk of breast cancer. There is an approximate 5% decrease in risk with each year delay in the initiation of menstruation [35]. It has been postulated that younger age at menarche corresponds to longer cumulative hormonal exposure and therefore elevated levels of estradiol [3, 36]. SNP rs4770552 is an intronic variant within the spermatogenesis associated 13 gene (*SPATA13*) at 13q12. *SPATA13* encodes a guanine nucleotide exchange factor (GEF) for RhoA, Rac1 and CDC42 GTPases [37, 38]. Although the role of this gene in breast cancer is still unclear, there could be an indirect link via the role of RhoA GTPases in breast tumorigenesis. Rho GTPase signaling is altered in human breast cancers, and dysregulation of Rho GTPase may have differential effects on the development of breast tumors depending on the stage and subtype [39]. Activation of RhoA results in release of megakaryoblastic leukemia 1 (*MKLI*), which in turn has been observed to alter the transcriptional activity of ER $\alpha$ , known to play a critical role in breast tumors [40]. Therefore, SNP rs4770552 may potentially indirectly interact with the regulatory region of *SPATA13* and affect the breast tumorigenesis process via activation of RHOA GTPases.

Given that the marginal effects of the common genetic variants are small and the associations of environmental risk factors with breast cancer are modest, interactions are also expected to be weak (Additional file 1: Figure S5). Although this is the largest breast cancer dataset available to date with more than 60,000 cases and 70,000 controls, the study is underpowered to detect weak interactions. Also, this study included only women of European ancestry and the findings may not be generalizable to women of other ancestries.

Using LDSC regression, we estimated the overall heritability due to G $\times$ E for each of the risk factors. The estimated frailty scale heritability ( $\leq 0.015$ ) can be compared with corresponding heritability for the SNP main effects (for which heritability is about 0.47) or the overall heritability based on the familial risk ( $\sim 1.4$ ) [28, 41]. The implication is that G $\times$ E interactions make very little contribution to the heritability of breast cancer, at least for the known risk factors and common genetic variants that can be evaluated using genome-wide arrays, and hence do not make an important contribution to risk prediction at the population level. This is consistent with the fact that detection of G $\times$ E interactions is rare. This does not rule out the possibility that G $\times$ E interactions could be identified in additional large studies or that such interactions may provide important clues to mechanisms.



## Conclusions

In conclusion, we identified two novel genome-wide gene–environment interactions for overall and ER+ breast cancer risk for women of European ancestry. These results contribute to our global body of knowledge on genetic susceptibility for breast cancer by generating plausible biological hypotheses, but they require replication and further functional studies.

## Abbreviations

BMI	Body mass index
COGS	Collaborative Oncological Gene-Environment Study
SNPs	Single nucleotide polymorphisms
GWAS	Genome-wide association study
GEWIS	Genome-wide gene–environment interaction study
G×E	Gene–environment interaction
BCAC	Breast Cancer Association Consortium
ER+	Estrogen receptor positive
ER-	Estrogen receptor negative
Q–Q	Quantile–quantile
BFDP	Bayesian false discovery probability
MAF	Minor allele frequency
OR <sub>int</sub>	Interaction odds ratio
CI	Confidence intervals
OR <sub>meta</sub>	Meta-analyzed odds ratio
LD	Linkage disequilibrium
IGF1	Insulin-like growth factor 1
LINGO2	Leucine rich repeat and Ig domain containing 2
SPATA13	Spermatogenesis associated 13
MKL1	Megakaryoblastic leukemia 1
LDSC	Linkage disequilibrium score regression

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13058-023-01691-8>.

**Additional file 1.** A genome-wide gene–environment interaction study of breast cancer risk for women of European ancestry. **Supplementary Table 1:** Participating studies with number of total cases and controls per study. **Supplementary Table 2:** Detailed information of the characteristics of the study population by study design and case-control status. **Supplementary Table 3:** Associations of epidemiological risk factors for overall and ER-specific subtype breast cancer risk in population-based and cohort studies. **Supplementary Table 4:** Stratified analysis results for genome-wide significant interaction results by categories of risk factors. **Supplementary Figure 1:** Quantile–Quantile (Q–Q) plots of genome-wide interaction of A) Adult height on overall breast cancer risk and B) Age at menarche on ER+ breast cancer risk. **Supplementary Figure 2:** Frailty-scale heritability explained by G×E interaction on overall and estrogen receptor positive breast cancer risk. **Supplementary Figure 3:** Regional association plot for the interaction analyses between SNP rs80018847 and adult height for overall breast cancer risk. **Supplementary Figure 4:** Regional association plot for the interaction analyses between SNP rs4770552 and age at menarche for ER+ breast cancer risk. **Supplementary Figure 5:** Power (x-axis) to detect gene–environment interaction odds ratio (y-axis) at different minor allele frequencies (0.01 to 0.5; legend below) for 1:1 unmatched case-control study for different sample sizes (N = 40,000 to 120,000 with 10,000 increment). Power calculation was performed by Quanto 1.2.4, assuming a log additive model with SNP marginal effect estimate as 1.10, marginal effect estimate of the environmental risk factor as 1.20, and a two-side alpha of  $5 \times 10^{-8}$ . We also assumed a 15% prevalence of the environmental risk factor and 1% prevalence of the disease.

## Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Deborah Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirmala Pathmanathan, Peter Simpson, J. Dinny Graham, Mythily Sachchithanathan. Samples are made available to researchers on a non-exclusive basis. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, Rachel Lloyd. The BCINIS study would not have been possible without the major contribution of Ms. H. Rennert, and the contributions of Dr. M. Pinchev, Dr. O. Barnet, Dr. N. Gronich, Dr. K. Landsman, Dr. A. Flugelman, Dr. W. Saliba, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. The BREGAN study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castela, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, María Elena Martínez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, María Torres, Manuel Calaza (BREGAN), 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-Carrero and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigación Biomedica Galicia Sur, SERGAS, Vigo, Spain. CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. CGPS thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadóttir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. CNIO-BCS thanks Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO). Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. The authors would like to thank the California Teachers Study Steering Committee that is responsible for the formation and maintenance of the Study within which this research was conducted. A full list of California Teachers Study (CTS) team members is available at <https://www.calteachersstudy.org/team>. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach. PROCAS thank NIHR for funding. The GENICA Network: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [IHB, RH, Wing-Yee Lo], Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany [YDK, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [UH], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. KARMA and SASBAC thank the Swedish Medical Research Council. kConFab/AOCS wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. MARIE thanks Petra Seibold, Nadia Obi, Ursula Eilber and Muhabbet Celik. The MCCS was made possible by the contribution

of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. The MISS study group acknowledges the former Principal Investigator, professor Håkan Olsson. NBHS and SBCGS thank study participants and research staff for their contributions and commitment to the studies. For NHS and NHS2 the study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We would like to thank the participants and staff of the NHS and NHS2 for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. OBCS thanks Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri Otsukka, Leena Keskitalo and Kari Mononen for their contributions to this study. The OFBCR thanks Teresa Selander, Nayana Weerasooriya and Steve Gallinger. ORIGO thanks E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. PBCS thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. We thank the SEARCH and EPIC teams. UCIBCS thanks Irene Masunaka. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible.

#### Author contributions

PM: Conceptualization, formal analysis, investigation, methodology, writing-review and editing. M.G.C.: Data curation, writing-review and editing. M.K.S.: Data curation, writing-review and editing. PK.: Data curation, writing-review and editing. R.L.M.: Data curation, writing-review and editing. S.L.: Data curation, writing-review and editing. D.F.E.: Resources, data curation, supervision, methodology, writing-review and editing. J.C.C.: Conceptualization, resources, supervision, methodology, writing-review and editing. All authors contributed in data curation and review of the manuscript.

#### Funding

Open Access funding enabled and organized by Projekt DEAL. BCAC is funded by the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Additional funding for BCAC is provided via the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer Research UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for iCOGS came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (U19 CA148537, U19 CA148065 and U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen

Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The BRIDGES panel sequencing was supported by the European Union Horizon 2020 research and innovation program BRIDGES (grant number, 634935) and the Wellcome Trust (v203477/Z/16/Z). The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society [grants NKI 2007–3839; 2009 4363]. The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute (grant number Z01-CP010119), and the National Institute of Environmental Health Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BCEES was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). For the BCFR-NY, BCFR-PA, BCFR-UT this work was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. The BCINIS study is supported in part by the Breast Cancer Research Foundation (BCRF). The BREast Oncology GALician Network (BREGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado and FEDER PI17/00918/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigación Biomedica Galicia Sur. Xerencia de Xestión Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economía y Competitividad, Xunta de Galicia, Spain. CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. CCGP is supported by funding from the University of Crete. The CECILE study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The CNIO-BCS was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The California Teachers Study (CTS) and the research reported in this publication were supported by the National Cancer Institute of the National Institutes of Health under award number U01-CA199277; P30-CA033572; P30-CA023100; UM1-CA164917; and R01-CA077398. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. The collection of cancer incidence data used in the California Teachers Study was supported by the California Department of Public Health pursuant to California Health and Safety Code Sect. 103885; Centers for Disease Control and Prevention's National Program of Cancer Registries, under cooperative agreement 5NU58DP006344; the National

Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201800032I awarded to the University of California, San Francisco, contract HHSN261201800015I awarded to the University of Southern California, and contract HHSN261201800009I awarded to the Public Health Institute. The opinions, findings, and conclusions expressed herein are those of the author(s) and do not necessarily reflect the official views of the State of California, Department of Public Health, the National Cancer Institute, the National Institutes of Health, the Centers for Disease Control and Prevention or their Contractors and Subcontractors, or the Regents of the University of California, or any of its programs. The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Cancer Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), P113/00061 to Granada, P113/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). PROCAS thank NIHR for funding. The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany. The GESBC was supported by the Deutsche Krebshilfe e.V. [70492] and the German Cancer Research Center (DKFZ). The KARMA study was supported by Märit and Hans Rausing's Initiative Against Breast Cancer. The KBPC was financially supported by the special Government Funding (VTR) 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. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, 199600). G.C.T. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The MCBCS was supported by the NIH grants R35CA253187, R01CA192393, R01CA116167, R01CA176785 a NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [P50CA116201], and the Breast Cancer Research Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index

and the Australian Cancer Database. The MEC was supported by NIH Grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study was supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society CAN 2018/675, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation FBKS 2021-19, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. The NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grant U01CA164920 from the USA National Cancer Institute of the National Institutes of Health. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The Carolina Breast Cancer Study (NCBCS) was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. The SASBAC study was supported by funding from the Agency for Science, Technology and Research of Singapore (A\*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank. SEARCH is funded by Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01.

#### Availability of data and materials

The datasets analyzed during the current study are not publicly available but are available upon request and approval of BCAC Data Access Co-ordinating Committee.

#### Declarations

##### Ethics approval and consent to participate.

Each participating study obtained informed consent from the participants and was approved by their local ethics committee.

##### Consent for participate

Each participating study obtained informed consent from the participants to publish and was approved by their local ethics committee.

##### Competing Interests

The authors do not have any conflict of interest.

**Author details**

<sup>1</sup>Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>2</sup>Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA. <sup>3</sup>Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. <sup>4</sup>Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. <sup>5</sup>Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. <sup>6</sup>Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. <sup>7</sup>Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada. <sup>8</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada. <sup>9</sup>Department of Medicine, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA. <sup>10</sup>Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>11</sup>Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada. <sup>12</sup>Division of Biostatistics, Institute for Health and Equity, and Cancer Center, Medical College of Wisconsin, Milwaukee, WI, USA. <sup>13</sup>Oncology, Clinical Sciences in Lund, Lund University, Lund, Sweden. <sup>14</sup>Department of Oncology, Leuven Multidisciplinary Breast Center, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium. <sup>15</sup>Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>16</sup>Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. <sup>17</sup>Human Genetics Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. <sup>18</sup>Centre for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain. <sup>19</sup>Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. <sup>20</sup>Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. <sup>21</sup>Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>22</sup>Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany. <sup>23</sup>iFIT-Cluster of Excellence, University of Tübingen, Tübingen, Germany. <sup>24</sup>German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Tübingen, Tübingen, Germany. <sup>25</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. <sup>26</sup>German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>27</sup>Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada. <sup>28</sup>Department of Biology, University of Pisa, Pisa, Italy. <sup>29</sup>Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>30</sup>Genomic Medicine Group, International Cancer Genetics and Epidemiology Group, Fundación Pública Galega de Medicina Xenómica, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain. <sup>31</sup>Grupo de Medicina Xenómica, Centro de Investigación en Red de Enfermedades Raras (CIBERER) y Centro Nacional de Genotipado (CEGEN-PRB2), Universidad de Santiago de Compostela, Santiago de Compostela, Spain. <sup>32</sup>Oncology and Genetics Unit, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestión Integrada de Vigo-SERGAS, Vigo, Spain. <sup>33</sup>Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. <sup>34</sup>Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA, USA. <sup>35</sup>City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA, USA. <sup>36</sup>Team 'Exposome and Heredity', CESP, Gustave Roussy, INSERM, University Paris-Saclay, UVSQ, Villejuif, France. <sup>37</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. <sup>38</sup>Department of Oncology and Metabolism, Sheffield Institute for Nucleic Acids (SiNFoNiA), University of Sheffield, Sheffield, UK. <sup>39</sup>Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK. <sup>40</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>41</sup>Nutrition and Metabolism Section, International Agency for Research on Cancer (IARC-WHO), Lyon, France. <sup>42</sup>Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia. <sup>43</sup>Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia. <sup>44</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

<sup>45</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>46</sup>Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>47</sup>Division of Evolution and Genomic Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, School of Biological Sciences, University of Manchester, Manchester, UK. <sup>48</sup>North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, Manchester Academic Health Science Centre, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, UK. <sup>49</sup>Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK. <sup>50</sup>Cancer Research UK Edinburgh Centre, The University of Edinburgh, Edinburgh, UK. <sup>51</sup>The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK. <sup>52</sup>Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. <sup>53</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia. <sup>54</sup>Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. <sup>55</sup>Institute for Clinical Research and Systems Medicine, Health and Medical University, Potsdam, Germany. <sup>56</sup>Department of Public Health Sciences, Queen's University, Kingston, ON, Canada. <sup>57</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>58</sup>Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. <sup>59</sup>Department of Oncology, Södersjukhuset, Stockholm, Sweden. <sup>60</sup>Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>61</sup>Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, Amherst, MA, USA. <sup>62</sup>Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK. <sup>63</sup>Nightingale and Genesis Prevention Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK. <sup>64</sup>NIHR Manchester Biomedical Research Unit, Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust, Manchester, UK. <sup>65</sup>Saarland Cancer Registry, Saarbrücken, Germany. <sup>66</sup>University of Tübingen, Tübingen, Germany. <sup>67</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. <sup>68</sup>Division of Cancer Sciences, University of Manchester, Manchester, UK. <sup>69</sup>Nuffield Department of Population Health, University of Oxford, Oxford, UK. <sup>70</sup>Surgery, Clinical Sciences in Lund, Lund University, Lund, Sweden. <sup>71</sup>Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, NSW, Australia. <sup>72</sup>Research Department, Peter MacCallum Cancer Center, Melbourne, VIC, Australia. <sup>73</sup>Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, VIC, Australia. <sup>74</sup>Department of Surgery, Kristianstad Hospital, Kristianstad, Sweden. <sup>75</sup>Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA. <sup>76</sup>Division of Oncology, Department of Medicine, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. <sup>77</sup>Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands. <sup>78</sup>Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. <sup>79</sup>Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany. <sup>80</sup>Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium. <sup>81</sup>VIB Center for Cancer Biology, VIB, Leuven, Belgium. <sup>82</sup>Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA. <sup>83</sup>Department of Surgical Sciences, Uppsala University, Uppsala, Sweden. <sup>84</sup>Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA. <sup>85</sup>Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel. <sup>86</sup>Department of Cancer Epidemiology and Prevention, M. Skłodowska-Curie National Research Oncology Institute, Warsaw, Poland. <sup>87</sup>Moore's Cancer Center and Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, USA. <sup>88</sup>Cancer Epidemiology Group, University Cancer Center Hamburg (UCC), University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>89</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada. <sup>90</sup>Laboratory Medicine Program, University Health Network, Toronto, ON, Canada. <sup>91</sup>INSERM UMR-S1138. CRB EPIGENETEC, Université Paris Cité, Paris, France. <sup>92</sup>School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada. <sup>93</sup>Cancer Control Research, BC Cancer, Vancouver, BC, Canada. <sup>94</sup>Epidemiology Branch, National Institute

of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA. <sup>95</sup>Department of Population Science, American Cancer Society, Atlanta, GA, USA. <sup>96</sup>Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. <sup>97</sup>Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. <sup>98</sup>Department of Oncology, Mayo Clinic, Rochester, MN, USA. <sup>99</sup>Department of Quantitative Health Sciences, Division of Clinical Trials and Biostatistics, Mayo Clinic, Rochester, MN, USA. <sup>100</sup>Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA. <sup>101</sup>Department of Surgical Oncology, University Hospitals Leuven, Leuven, Belgium. <sup>102</sup>Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC, Australia. <sup>103</sup>Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, WA, Australia. <sup>104</sup>Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA. <sup>105</sup>Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA. <sup>106</sup>Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>107</sup>Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA. <sup>108</sup>Division of Computational Biology, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA. <sup>109</sup>Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid, Spain. <sup>110</sup>Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni Van Leeuwenhoek Hospital, Amsterdam, the Netherlands. <sup>111</sup>Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

Received: 5 April 2023 Accepted: 27 July 2023

Published online: 09 August 2023

## References

- Baer HJ, et al. Adult height, age at attained height, and incidence of breast cancer in premenopausal women. *Int J Cancer*. 2006;119:2231–5.
- Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the disease. *Lancet*. 2002;360:187–95.
- Collaborative Group on Hormonal Factors in Breast Cancer. 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.
- Hunter DJ, et al. Oral contraceptive use and breast cancer: a prospective study of young women. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol*. 2010;19:2496–502.
- Jung S, et al. Alcohol consumption and breast cancer risk by estrogen receptor status: in a pooled analysis of 20 studies. *Int J Epidemiol*. 2016;45:916–28.
- Wang K, et al. Change in risk of breast cancer after receiving hormone replacement therapy by considering effect-modifiers: a systematic review and dose-response meta-analysis of prospective studies. *Oncotarget*. 2017;8:81109–24.
- World Cancer Research Fund International/American Institute for Cancer Research. Diet, nutrition, physical activity and breast cancer 2017. 2017;120.
- Amos CI, et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomark Prev*. 2017;26:126–35.
- Michailidou K, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551:92–4.
- Milne RL, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet*. 2017;49:1767–78.
- Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*.
- Fletcher O, Dudbridge F. Candidate gene-environment interactions in breast cancer. *BMC Med*. 2014;12:195.
- Hutter CM, Mechanic LE, Chatterjee N, Kraft P, Gillanders EM. Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genet Epidemiol*. 2013;37:643–57.
- Rudolph A, Chang-Claude J, Schmidt MK. Gene-environment interaction and risk of breast cancer. *Br J Cancer*. 2016;114:125–33.
- Simonds NJ, et al. Review of the gene-environment interaction literature in cancer: What do we know? *Genet Epidemiol*. 2016;40:356–65.
- Hein R, et al. A genome-wide association study to identify genetic susceptibility loci that modify ductal and lobular postmenopausal breast cancer risk associated with menopausal hormone therapy use: a two-stage design with replication. *Breast Cancer Res Treat*. 2013;138:529–42.
- Rudolph A, et al. Genetic modifiers of menopausal hormone replacement therapy and breast cancer risk: a genome-wide interaction study. *Endocr Relat Cancer*. 2013;20:875–87.
- Wang X, et al. Genome-wide interaction analysis of menopausal hormone therapy use and breast cancer risk among 62,370 women. *Sci Rep*. 2022;12:6199.
- Schoeps A, et al. Identification of new genetic susceptibility loci for breast cancer through consideration of gene-environment interactions. *Genet Epidemiol*. 2014;38:84–93.
- Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013;45:353–361e2.
- Michailidou K, 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.
- Amos CI, et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomark Prev*. 2017;26:126–35.
- Fachal L, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat Genet*. 2020;52:56–73.
- Kapoor PM, et al. Assessment of interactions between 205 breast cancer susceptibility loci and 13 established risk factors in relation to breast cancer risk, in the Breast Cancer Association Consortium. *Int J Epidemiol*. 2020;49:216–32.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–1.
- Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet*. 2007;81:208–27.
- Bulik-Sullivan BK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–5.
- Michailidou K, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*.
- Friedenreich CM. Review of anthropometric factors and breast cancer risk. *Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP*. 2001;10:15–32.
- Gunnell D, et al. Height, leg length, and cancer risk: a systematic review. *Epidemiol Rev*. 2001;23:313–42.
- Zhang B, et al. Height and breast cancer risk: evidence from prospective studies and Mendelian randomization. *JNCI J Natl Cancer Inst*. 2015;107:djv219.
- Nunney L. Size matters: height, cell number and a person's risk of cancer. *Proc R Soc B Biol Sci*. 2018;285:20181743.
- Renahan AG. Height and cancer: consistent links, but mechanisms unclear. *Lancet Oncol*. 2011;12:716–7.
- Haines BP, Rigby PWJ. Expression of the Lingo/LERN gene family during mouse embryogenesis. *Gene Expr Patterns*. 2008;8:79–86.
- Fortner R, Hankinson SE. Reproductive and hormonal factors and breast cancer. 2012. <https://doi.org/10.1210/TEAM.9781936704064.CH4>.
- Endogenous Hormones and Breast Cancer Collaborative Group, et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. *Lancet Oncol*. 2013;14:1009–19.
- Kawasaki Y, et al. Identification and characterization of Asef2, a guanine-nucleotide exchange factor specific for Rac1 and Cdc42. *Oncogene*. 2007;26:7620–7.
- Bristow JM, et al. The Rho-family GEF Asef2 activates Rac to modulate adhesion and actin dynamics and thereby regulate cell migration. *J Cell Sci*. 2009;122:4535–46.

39. McHenry PR, Vargo-Gogola T. Pleiotropic functions of Rho GTPase signaling: A Trojan horse or Achilles' heel for breast cancer treatment? *Curr Drug Targets*. 2010;11:1043–58.
40. Huet G, et al. Repression of the estrogen receptor- $\alpha$  transcriptional activity by the Rho/megakaryoblastic leukemia 1 signaling pathway. *J Biol Chem*. 2009;284:33729–39.
41. Pharoah PDP, et al. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet*. 2002;31:33–6.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

