Genetic variation at CYP3A is associated with age at menarche and breast cancer risk: a case-control study


Published in:
Breast Cancer Research

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright the authors 2014.
This is an open access article published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen’s institutional repository that provides access to Queen’s research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Download date:30. Sep. 2019
Genetic variation at CYP3A is associated with age at menarche and breast cancer risk: a case-control study


© 2014 Johnson et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

http://breast-cancer-research.com/content/16/3/R51
Chen-Yang Shen12,13, Jyh-Cheng Yu12,4, Chiu-Sheng Huang12,5, Ming-Feng Hou12,6,17, Anna González-Neira29, Daniel C Tessier12,6, Daniel Vincent12,8, Francois Bacot12,8, Craig Luccarini102, Joe Dennis129, Kyriaki Michailidou129, Manjeet K Bolla12,9, Jean Wang12,9, Douglas F Easton102,129, Montserrat García-Closas1,2,4, Mitch Dowsett5, Alan Ashworth1,2, Anthony J Swerdlow12,1,2, Julian Peto3, Isabel dos Santos Silva31 and Olivia Fletcher1,2

Abstract

Introduction: We have previously shown that a tag single nucleotide polymorphism (rs10235235), which maps to the CYP3A locus (7q22.1), was associated with a reduction in premenopausal urinary estrone glucuronide levels and a modest reduction in risk of breast cancer in women age ≤50 years.

Methods: We further investigated the association of rs10235235 with breast cancer risk in a large case control study of 47,346 cases and 47,570 controls from 52 studies participating in the Breast Cancer Association Consortium. Genotyping of rs10235235 was conducted using a custom Illumina Infinium array. Stratified analyses were conducted to determine whether this association was modified by age at diagnosis, ethnicity, age at menarche or tumor characteristics.

Results: We confirmed the association of rs10235235 with breast cancer risk for women of European ancestry but found no evidence that this association differed with age at diagnosis. Heterozygote and homozygote odds ratios (ORs) were OR = 0.98 (95% CI 0.94, 1.01; P = 0.2) and OR = 0.80 (95% CI 0.69, 0.93; P = 0.004), respectively (P_{trend} = 0.02). There was no evidence of effect modification by tumor characteristics. rs10235235 was, however, associated with age at menarche in controls (P_{trend} = 0.005) but not cases (P_{trend} = 0.97). Consequently the association between rs10235235 and breast cancer risk differed according to age at menarche (P_{het} = 0.02); the rare allele of rs10235235 was associated with a reduction in breast cancer risk for women who had their menarche age ≥15 years (OR_{het} = 0.84, 95% CI 0.75, 0.94; OR_{norm} = 0.81, 95% CI 0.51, 1.30; P_{trend} = 0.002) but not for those who had their menarche age ≤11 years (OR_{het} = 1.06, 95% CI 0.81, 1.39, OR_{norm} = 1.07, 95% CI 0.67, 1.72; P_{trend} = 0.29).

Conclusions: To our knowledge rs10235235 is the first single nucleotide polymorphism to be associated with both breast cancer risk and age at menarche consistent with the well-documented association between later age at menarche and a reduction in breast cancer risk. These associations are likely mediated via an effect on circulating hormone levels.

Introduction

Family history is a well-established risk factor for breast cancer. First-degree relatives of women with breast cancer have an approximately twofold increased risk of developing the disease relative to the general population [1]. Twin studies are consistent with this familial clustering having, at least in part, a genetic origin [2,3]. Mutations in high-risk susceptibility genes (mainly BRCA1 and BRCA2) explain most large multiple-case families, but account for only 15 to 20% of the excess familial risk [4]. Genome-wide association studies [5,6] have identified more than 70 common variants that are associated with breast cancer susceptibility but they account for only another approximately 15% of the excess familial risk. The so-called ‘missing heritability’ may be explained by common variants with very small effects and/or by rarer variants with larger effects, neither of which can be identified by current genome-wide association studies. A statistically efficient alternative is to increase power by trying to identify variants associated with known quantitative phenotypic markers of susceptibility to breast cancer [7], and then to test them for association with breast cancer risk. This approach might also improve our understanding of the biological mechanisms involved in breast cancer pathogenesis.

Endogenous sex hormones are well-established risk factors for breast cancer in postmenopausal women [8]; the evidence in premenopausal women is less consistent, with some, but not all, studies suggesting an association between higher circulating levels of estrogens and increased breast cancer risk [9-17]. Genetic factors influence the levels of endogenous sex hormones [18] and therefore single nucleotide polymorphisms (SNPs) in genes regulating these hormonal pathways are good candidates for being breast cancer predisposition variants. We have previously studied 642 SNPs tagging 42 genes that might influence sex hormone levels in 729 healthy premenopausal women of European ancestry in relation to cyclic variations in oestrogen levels during the menstrual cycle. We found that the minor allele of rs10273424, which maps 50 kb 3’ to CYP3A4, was associated with a reduction of 22% (95% confidence interval (CI) = −28%, −15%; P = 10^{-19}) in levels of urinary oestrone glucuronide, a metabolite that is highly correlated with serum oestradiol levels [19]. Analysis of 10,551 breast cancer cases and 17,535 controls of European ancestry demonstrated that the minor allele of rs10235235, a proxy for rs10273424 (r^2 = 1.0), was also associated with a weak reduction in...
breast cancer risk but only in women aged 50 years or younger at diagnosis (odds ratio (OR) = 0.91, 95% CI = 0.83, 0.99; \( P = 0.03 \)) [19].

The aim of the present study was to further investigate an association between rs10235235 and breast cancer risk using a much larger set of subjects – the Breast Cancer Association Consortium (BCAC) – comprising data from 49 additional studies, and to assess whether there was evidence of effect modification by age at diagnosis, ethnicity, age at menarche or tumour characteristics.

Materials and methods

Sample selection

Samples for the case–control analyses were drawn from 52 studies participating in the BCAC: 41 studies from populations of predominantly European ancestry, nine studies of Asian ancestry and two studies of African-American ancestry. The majority were population-based or hospital-based case–control studies, but some studies were nested in cohorts, selected samples by age, oversampled for cases with a family history or selected samples on the basis of tumour characteristics (Table S1 in Additional file 1). Studies provided ~2% of samples in duplicate for quality control purposes (see below). Study subjects were recruited on protocols approved by the Institutional Review Boards at each participating institution, and all subjects provided written informed consent (Additional file 2).

Genotyping and post-genotyping quality control

Genotyping for rs10235235 was carried out as part of a collaboration between the BCAC and three other consortia (the Collaborative Oncological Gene-environment Study (COGS)). Full details of SNP selection, array design, genotyping and post-genotyping quality control have been published [5]. Briefly, three categories of SNPs were chosen for inclusion in the array: SNPs selected on the basis of pooled genome-wide association study data; SNPs selected for the fine-mapping of published risk loci; and candidate SNPs selected on the basis of previous analyses or specific hypotheses. rs10235235 was a candidate SNP selected on the basis of our previous analyses [19].

For the COGS project overall, genotyping of 211,155 SNPs in 114,225 samples was conducted using a custom Illumina Infinium array (iCOGS; Illumina, San Diego, CA, USA) in four centres. Genotypes were called using Illumina’s proprietary GenCall algorithm. Standard quality control measures were applied across all SNPs and all samples genotyped as part of the COGS project. Samples were excluded for any of the following reasons: genotypically not female XX (XY, XXY or XO, \( n = 298 \)); overall call rate <95% (\( n = 1,656 \)); low or high heterozygosity (\( P < 10^{-6} \), separately for individuals of European, Asian and African-American ancestry, \( n = 670 \)); individuals not concordant with previous genotyping within the BCAC (\( n = 702 \)); individuals where genotypes for the duplicate sample appeared to be from a different individual (\( n = 42 \)); cryptic duplicates within studies where the phenotypic data indicated that the individuals were different, or between studies where genotype data indicated samples were duplicates (\( n = 485 \)); first-degree relatives (\( n = 1,981 \)); phenotypic exclusions (\( n = 527 \)); or concordant replicates (\( n = 2,629 \)).

Ethnic outliers were identified by multidimensional scaling, combining the iCOGS array data with the three Hapmap2 populations, based on a subset of 37,000 uncorrelated markers that passed quality control (including ~1,000 selected as ancestry informative markers). Most studies were predominantly of a single ancestry (European or Asian), and women with >15% minority ancestry, based on the first two components, were excluded (\( n = 1,244 \)). Two studies from Singapore (SGBCC) and Malaysia (MYBRCa; see Table S1 in Additional file 1 for all full study names) contained a substantial fraction of women of mixed European/Asian ancestry (probably of South Asian ancestry). For these studies, no exclusions for ethnic outliers were made, but principal components analysis (see below) was used to adjust for inflation in these studies. Similarly, for the two African-American studies (NBHS and SCCS), no exclusions for ethnic outliers were made.

Principal component analyses were carried out separately for the European, Asian and African-American subgroups, based on a subset of 37,000 uncorrelated SNPs. For the analyses of European subjects, we included the first six principal components as covariates, together with a seventh component derived specific to one study (LMBC) for which there was substantial inflation not accounted for by the components derived from the analysis of all studies. Addition of further principal components did not reduce inflation further. Two principal components were included for the studies conducted in Asian populations and two principal components were included for the African-American studies.

For the main analyses of rs10235235 and breast cancer risk, we excluded women from three studies (BBCS, BIGGS and UKBGS) that were genotyped in the hypothesis-generating study (\( n = 5,452 \)) [19] and women with non-invasive cancers (ductal carcinoma in situ/lobular carcinoma in situ, \( n = 2,663 \) or cancers of uncertain status (\( n = 960 \)). After exclusions there were 47,346 invasive breast cancer case samples and 47,570 control samples from 49 studies (38 from populations of predominantly European ancestry, nine Asian and two African-American) used in the analysis (Tables S1 and S2 in Additional file 1). After quality control exclusions (above) the call rate for rs10235235 was 100% (one no call in 94,916 samples), and for the controls there was no evidence of deviation from
Hardy–Weinberg equilibrium in any of the contributing studies (Table S2 in Additional file 1).

We did not test for an association between rs10235235 and age at menarche in our hypothesis-generating study [19]. Therefore, to maximise our power to detect an association, we included menarche data from BBCS cases (n = 2,508) and controls (n = 1,650) and from UKBGS cases (n = 3,388) and controls (n = 4,081) in this analysis. Age at menarche was not available for samples from BIGGS. Full details of genotyping of rs10235235 in BBCS and UKBGS samples have been published previously [19]. Briefly, genotyping was carried out using competitive allele-specific polymerase chain reaction KASPar chemistry (KBiosciences Ltd, Hoddesdon, Hertfordshire, UK). Call rates were 98.0% (BBCS) and 96.6% (UKBGS); there was no evidence for deviation from Hardy–Weinberg equilibrium (P = 0.29 (BBCS); P = 0.92 (UKBGS)), and the duplicate concordance based on a 1% (BBCS) and 5% (UKBGS) random sample of duplicates was 100% for both studies.

Statistical analysis

We estimated per-allele and genotypic log odds ratios (ORs) for the European, Asian and African-American subgroups separately using logistic regression, adjusted for principal components and study [5]. To test for departure from a multiplicative model we compared multiplicative and unconstrained models using a one degree of freedom test with one degree of freedom. To test for effect modification of an association between age at menarche and breast cancer risk. To test for effect modification of an association between rs10235235 and breast cancer risk by age at menarche, we used logistic regression adjusted for principal components, study and age at menarche (grouped as ≤11, 12, 13, 14 and ≥15 years) with and without an interaction term(s). We considered four models: no interaction (zero interaction terms); assuming a linear interaction between genotype and menarche group (one interaction term); assuming a linear interaction between genotype and menarche group but allowing the linear term to differ between women who were heterozygous and those who were homozygous for the rare allele (two interaction terms); and one interaction term for each possible genotype/menarche group combination (eight interaction terms). Nested models were compared using likelihood ratio tests. All statistical analyses were performed using STATA version 11.0 (StataCorp, College Station, TX, USA). All P values reported are two-sided.

Results

The case–control analysis comprised genotype data for 47,346 invasive breast cancer cases and 47,569 controls from 49 studies, including 80,518 (84.8%) subjects of self-reported European ancestry, 12,419 (13.1%) of self-reported Asian ancestry and 1,978 (2.1%) of self-reported African-American ancestry. The mean (± standard deviation) age at diagnosis was 56.1 (± 11.6) years for European cases, 51.1 (± 10.5) years for Asian cases and 53.1 (± 10.7) years for African-American cases. There were ethnic differences in the estimated minor allele frequency (MAF) of rs10235235 (Q = 7317.1, two degrees of freedom; P for heterogeneity (P_{het}) = 0). The overall MAF for European control women was 0.089 (95% CI = 0.087, 0.091), but with strong evidence of between-study heterogeneity (P_{het} = 1 × 10^{−22}) that was accounted for by the three Finnish studies (HEBCS, MAF = 0.15; KBCP, MAF = 0.21; and OBBCS, MAF = 0.15; P_{het} = 0.01); no evidence of heterogeneity remained after taking account of these studies (MAF = 0.087 (95% CI = 0.085, 0.089); P_{het} = 0.23). Relative to Europeans, the overall MAF was higher for African-Americans (0.213, 95% CI = 0.195, 0.232; P_{het} = 0.26) but much lower for Asians (0.002; 95% CI = 0.001, 0.002), with strong evidence of between-study heterogeneity for the latter (P_{het} = 4 × 10^{−14}).
The case–control analysis was consistent with a modest association between rs10235235 and breast cancer risk for women of European ancestry, with an estimated per-allele OR of 0.96 (95% CI = 0.93, 0.99;  P  for linear trend ( P  trend) = 0.02). Genotype-specific ORs were 0.98 (95% CI = 0.94, 1.01;  P  = 0.21) for AG versus AA (Figure 1A) and 0.80 (95% CI = 0.69, 0.93;  P  = 0.004) for GG versus AA (Figure 1B), with no evidence of between-study heterogeneity for either OR estimate ( P  het = 0.44,  I 2 = 1.9% and  P  het = 0.76,  I 2 = 0.0% for heterozygote and homozygote OR estimates respectively). There was, however, marginally significant evidence that the genotypic OR estimates departed from those expected under a multiplicative model with the inverse association of the GG genotype being more than the square of that of the AG genotype (test for deviation from multiplicative model,  P  = 0.04).

Data for rs10235235 in women of Asian or African-American ancestry were more limited, with just two African-American studies (1,046 cases and 932 controls) and nine Asian studies (5,795 cases and 6,624 controls). In addition, this SNP was sufficiently rare in Asian populations (MAF = 0.002) that we were unable to estimate the heterozygote OR in two Asian studies (SEBCS, one carrier among 236 controls and no carriers among 1,129 controls; TWBCS, one carrier among 236 controls and no carriers among 774 cases; Table S2 in Additional file 1) and we could not estimate a homozygote OR for any Asian study (Table S2 in Additional file 1). There was no clear evidence that this SNP was associated with breast cancer risk for women of Asian ancestry (heterozygote OR = 1.06, 95% CI = 0.76, 1.49) or African-American ancestry (heterozygote and homozygote ORs were OR = 1.09, 95% CI = 0.90, 1.32 and OR = 0.94, 95% CI = 0.62, 1.42 respectively; Figure S1 in Additional file 1). This analysis, however, had low power to detect associations in non-Europeans and these OR estimates were not inconsistent with the magnitude of the observed OR estimates for European women ( P  het = 0.51).

Stratifying cases by oestrogen receptor ( P  het = 0.83) or progesterone receptor ( P  het = 0.19) status, tumour grade ( P  het = 0.63) or nodal involvement at diagnosis ( P  het = 0.51) showed no evidence of effect modification (Table 1). There was some evidence of effect modification by morphology ( P  het = 0.03). For ductal cancers we estimated a very modest reduction of risk for heterozygotes (OR het = 0.98, 95% CI = 0.93, 1.02;  P  = 0.30) and a stronger, significant reduction for homozygotes (OR nom = 0.74, 95% CI = 0.61, 0.90;  P  = 0.003). For lobular cancers there was no such trend (OR het = 1.07, 95% CI = 0.98, 1.17;  P  = 0.14 and OR nom = 0.91, 95% CI = 0.64, 1.27;  P  = 0.57).

The SNP rs10235235 maps to a locus (CYP3A7) that has been considered an a priori candidate for involvement in determining age at menopause and age at menarche [21,22]. Stratifying cases by age at diagnosis (≤50

---

**Figure 1** Association of rs10235235 with breast cancer risk for women of European ancestry. Forest plots of the association of the rs10235235 AG (heterozygote) genotype (A) and GG (homozygote) genotype (B) with breast cancer risk for women of European ancestry. Horizontal lines, 95% confidence intervals (CIs); square boxes, study-specific fixed-effects estimates; diamond, combined, fixed-effects estimate of the odds ratio (OR) and 95% CI. Vertical line, null effect (OR = 1.0); dashed vertical line, estimated heterozygote OR (A) and estimated homozygote OR (B). Homozygote ORs for six studies (CTS, DEMOKRITOS, kConFab/AOCS, NBCS, NBHS and RPCD) could not be estimated because there were no GG homozygotes among cases or among controls in each of these studies (see Table S2 in Additional file 1).
or >50 years) as a proxy for menopausal status at diagnosis showed no evidence of effect modification (P\textsubscript{het} = 0.89; Table 2), and excluding cases who were diagnosed between age 46 and 55 as potentially perimenopausal did not alter this result (P\textsubscript{het} = 0.28). Data on age at menarche were available for 21,736 cases and 22,686 controls (Table S4 in Additional file 1); to increase the power of the analysis we included additional data from BBCS and UKBGS (5,737 cases, 5,572 controls; Table S4 in Additional file 1) [19]. There was a 1.5% (95% CI = 0.5%, 2.7%; \(P = 0.004\)) reduction in breast cancer risk associated with each additional year’s increase in age at menarche. Mean age at menarche was positively associated with number of copies of the minor allele of rs10235235 for controls (\(P\text{trend} = 0.005\); Table 3) but not for cases (\(P\text{trend} = 0.97\); Table 3). Consequently, there was an inverse trend in the magnitude of the heterozygote and homozygote breast cancer ORs with mean age at menarche \(\geq 15\) years but there was no evidence of reduction for those with a menarche at age \(\leq 11\) years (OR\textsubscript{het} = 1.06, 95% CI = 0.95, 1.19; \(P = 0.30\) and OR\textsubscript{hom} = 1.07, 95% CI = 0.67, 1.72; \(P = 0.78\)) (\(P\text{trend} = 0.29\)). There was no evidence that the inverse trend in the magnitude of ORs with mean age at menarche differed between heterozygous and homozygous carriers (\(P = 0.97\)) and no evidence that the trend was nonlinear (\(P = 0.70\)).

### Table 1 Association of rs10235235 with risk of breast cancer for women of European ancestry: stratified analysis

<table>
<thead>
<tr>
<th>ER status</th>
<th>Cases</th>
<th>Controls</th>
<th>OR\textsubscript{het}</th>
<th>95% CI</th>
<th>(P_1)</th>
<th>OR\textsubscript{hom}</th>
<th>95% CI</th>
<th>(P_1)</th>
<th>(P_{het})</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER-positive</td>
<td>24,780</td>
<td>38,739</td>
<td>0.99</td>
<td>0.95, 1.03</td>
<td>0.61</td>
<td>0.83</td>
<td>0.70, 0.99</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>ER-negative</td>
<td>5,851</td>
<td>38,739</td>
<td>1.02</td>
<td>0.95, 1.10</td>
<td>0.60</td>
<td>0.60</td>
<td>0.43, 0.86</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>NK</td>
<td>8,339</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>38,970</td>
<td>38,739</td>
<td>0.99</td>
<td>0.95, 1.03</td>
<td>0.74</td>
<td>0.79</td>
<td>0.67, 0.94</td>
<td>0.006</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PR status</th>
<th>Cases</th>
<th>Controls</th>
<th>OR\textsubscript{het}</th>
<th>95% CI</th>
<th>(P_1)</th>
<th>OR\textsubscript{hom}</th>
<th>95% CI</th>
<th>(P_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR-positive</td>
<td>18,497</td>
<td>39,033</td>
<td>0.98</td>
<td>0.93, 1.02</td>
<td>0.32</td>
<td>0.82</td>
<td>0.67, 0.99</td>
<td>0.04</td>
</tr>
<tr>
<td>PR-negative</td>
<td>8,193</td>
<td>39,033</td>
<td>1.02</td>
<td>0.96, 1.09</td>
<td>0.53</td>
<td>0.74</td>
<td>0.56, 0.98</td>
<td>0.03</td>
</tr>
<tr>
<td>NK</td>
<td>12,111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>38,801</td>
<td>39,033</td>
<td>0.99</td>
<td>0.94, 1.03</td>
<td>0.52</td>
<td>0.80</td>
<td>0.67, 0.95</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Morphology</th>
<th>Cases</th>
<th>Controls</th>
<th>OR\textsubscript{het}</th>
<th>95% CI</th>
<th>(P_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ductal</td>
<td>22,123</td>
<td>31,803</td>
<td>0.98</td>
<td>0.93, 1.02</td>
<td>0.30</td>
</tr>
<tr>
<td>Lobular</td>
<td>3,921</td>
<td>31,803</td>
<td>1.07</td>
<td>0.98, 1.17</td>
<td>0.14</td>
</tr>
<tr>
<td>Other and NK</td>
<td>5,995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>32,039</td>
<td>31,803</td>
<td>0.99</td>
<td>0.95, 1.04</td>
<td>0.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>Cases</th>
<th>Controls</th>
<th>OR\textsubscript{het}</th>
<th>95% CI</th>
<th>(P_1)</th>
<th>OR\textsubscript{hom}</th>
<th>95% CI</th>
<th>(P_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1</td>
<td>5,944</td>
<td>37,285</td>
<td>0.97</td>
<td>0.90, 1.05</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 2</td>
<td>13,427</td>
<td>37,285</td>
<td>1.00</td>
<td>0.95, 1.06</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>8,638</td>
<td>37,285</td>
<td>0.98</td>
<td>0.92, 1.05</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NK</td>
<td>8,769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36,778</td>
<td>37,285</td>
<td>0.99</td>
<td>0.95, 1.03</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nodal status</th>
<th>Cases</th>
<th>Controls</th>
<th>OR\textsubscript{het}</th>
<th>95% CI</th>
<th>(P_1)</th>
<th>OR\textsubscript{hom}</th>
<th>95% CI</th>
<th>(P_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node-negative</td>
<td>17,463</td>
<td>37,836</td>
<td>0.98</td>
<td>0.93, 1.03</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Node-positive</td>
<td>10,746</td>
<td>37,836</td>
<td>0.98</td>
<td>0.92, 1.04</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NK</td>
<td>9,359</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>37,568</td>
<td>37,836</td>
<td>0.98</td>
<td>0.94, 1.02</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Association of rs10235235 with risk of breast cancer for women of European ancestry stratified by oestrogen receptor (ER) status, progesterone receptor (PR) status, morphology, grade and nodal status. OR\textsubscript{het}, odds ratio comparing rs10235235 AG genotype versus AA genotype; H0, null hypothesis; NK, not known; OR\textsubscript{hom}, odds ratio comparing rs10235235 GG genotype versus AA genotype; \(P_1\), test of H0 no association between rs10235235 and breast cancer risk; \(P_{het}\), test of H0 no difference between stratum specific estimates for variables with two strata or test of H0 no linear trend in stratum specific estimates for variables with three strata. aExcludes seven studies that selected all ER-negative cases (CTS, DEMOKRITOS, NBCS, NBHS, OSU, RPCI and SKKDKFZS) and one study (PBCS) that selected all ER-positive cases. bExcludes seven studies that selected all PR-negative cases (CTS, DEMOKRITOS, NBCS, NBHS, OSU, RPCI and SKKDKFZS).
Table 2 rs10235235 and risk of breast cancer for women of European ancestry by age at diagnosis

<table>
<thead>
<tr>
<th>Age at diagnosis</th>
<th>Cases*</th>
<th>Controls†</th>
<th>ORhet 95% CI</th>
<th>Pt</th>
<th>ORhom 95% CI</th>
<th>Pt</th>
<th>Phet</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 50 years</td>
<td>11,794</td>
<td>34,988</td>
<td>0.99</td>
<td>0.93, 1.05</td>
<td>0.69</td>
<td>0.08</td>
<td>0.53, 0.86</td>
</tr>
<tr>
<td>&gt; 50 years</td>
<td>23,264</td>
<td>34,988</td>
<td>0.97</td>
<td>0.93, 1.02</td>
<td>0.24</td>
<td>0.84</td>
<td>0.70, 1.00</td>
</tr>
<tr>
<td>NK</td>
<td>554</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35,612</td>
<td>34,988</td>
<td>0.98</td>
<td>0.94, 1.02</td>
<td>0.23</td>
<td>0.79</td>
<td>0.67, 0.92</td>
</tr>
</tbody>
</table>

*aFive studies (ABCFS, MARIE, MEC, MTU-GBCS and SABAC) that selected all cases on the basis of age at diagnosis (Table S3 in Additional file 1) were excluded from this stratified analysis; two small studies (CTS and NBCCS) that had no heterozygote or rare homozygote cases in one of the age stratum were also excluded. H0, null hypothesis; NK, not known; ORhet, odds ratio comparing rs10235235 AG genotype versus AA genotype; ORhom, odds ratio comparing rs10235235 GG genotype versus AA genotype; Pt, test of H0 no association between rs10235235 and breast cancer risk; Phet, test of H0 no difference between stratum specific estimates.

Discussion

This study of more than 47,000 breast cancer cases and 47,000 controls has confirmed that rs10235235, mapping to 7q22.1 (CYP3A4), is associated with a reduction in breast cancer risk for women of European ancestry. Previously, our hypothesis-generating study of 10,000 breast cancer cases and 17,000 controls found a per-allele OR estimate of 0.96 (95% CI = 0.90, 1.02; P = 0.2), with marginally significant evidence of an inverse association for breast cancer diagnosed age 50 years or younger (OR = 0.91, 95% CI = 0.83, 0.99; P = 0.03) but no evidence of an association for breast cancer at later ages (OR = 1.01, 95% CI = 0.93, 1.10; P = 0.82) [19]. In this considerably larger study, we found a heterozygote OR estimate of 0.98 (95% CI = 0.94, 1.01; P = 0.21) and a homozygote OR estimate of 0.80 (95% CI = 0.69, 0.93; P = 0.004) with marginally significant evidence that the inverse association for homozygotes is greater than predicted by a multiplicative model (P = 0.04).

To our knowledge, rs10235235 is the first SNP to be associated with both breast cancer risk and age at menarche, consistent with the well-documented association between later age at menarche and a reduction in breast cancer risk [23]. Genome-wide association studies have identified more than 70 breast cancer risk variants [5,6] and more than 30 variants associated with age at menarche [22], none of which map to the CYP3A4 locus. rs10235235 was originally identified on the basis of a highly significant association with hormone levels, accounting for 4.9% of the variation in premenopausal urinary oestrone glucuronide levels [19]. In this current analysis, rs10235235 accounted for only 0.01% of the variation across controls in age at menarche and we estimate that this SNP explains just 0.01% of the familial excess breast cancer risk. Our data thus illustrate the potential statistical efficiency of studies of intermediate phenotypes in the identification of rarer (MAF < 10%) risk alleles with modest associations. Our analysis shows some inconsistency with a recent genome-wide study of circulating oestradiol, testosterone and sex hormone-binding globulin in postmenopausal women [24]. In that study there was no genome-wide significant association observed with plasma oestradiol levels in either the primary analysis of approximately 1,600 postmenopausal women who were not taking postmenopausal hormones at blood draw or the secondary analysis that included approximately 900 current postmenopausal hormone users. Further studies will be needed to determine whether the lack of an association between CYP3A4 variants and postmenopausal plasma oestradiol levels reflects a difference in the menopausal status of the study subjects, the hormone/metabolite that was analysed or chance.

One possible explanation for the apparent effect modification of the rs10235235 breast cancer risk association by age at menarche is that this is a function of genotyping a marker SNP rather than the true causal variant. For example, if rs10235235 was perfectly correlated with a causal variant, SNP X, with a MAF substantially lower than that of rs10235235 (D’ ~ 1.0, r2 < 1.0), then there would be three types of chromosome in the population: type i, chromosomes carrying the common allele of rs10235235 and the common allele of SNP X; type ii, chromosomes carrying the rare allele of rs10235235 and the common allele of SNP X; and type iii, chromosomes carrying the rare allele of rs10235235 and the rare (protective) allele of SNP X. Only chromosomes carrying the rare allele of rs10235235 and the rare (protective) allele of...
SNP X (type iii) would be enriched in controls. Genotyping the marker (rs10235235) rather than the causal variant leads to misclassification. As the causal variant is associated with a protective effect on breast cancer risk, the proportion of chromosomes carrying both the rare allele of the causal variant and the marker (type iii) compared with the common allele of the causal variant and the rare allele of the marker (type ii) will be greater in controls than in cases such that the extent of misclassification will be greater for cases than controls. This will attenuate the association between genotype and age at menarche to a greater extent in cases than in controls creating an apparent effect modification. Fine mapping and functional studies will be required to identify the causal variant and to determine the true relationship between the causal variant, age at menarche and breast cancer risk.

Despite our original finding of a strong association between rs10235235 and hormone levels, we found no evidence that the association between this SNP and breast cancer risk differed by the hormone receptor status of the tumour, and nor did we find any evidence that the association differed by stage, grade or lymph node involvement. There was marginally significant evidence that the association between rs10235235 and breast cancer risk differed between ductal and lobular cancers ($P_{\text{het}} = 0.03$). Given the number of stratified analyses that we carried out (six stratifying variables) and given that there is no biological basis to support an interaction between rs10235235 and morphology, this is probably a chance observation.

In contrast to our earlier study [19], we found no evidence of an interaction with age at diagnosis when we stratified cases by age ≤/≥50 years, either including or excluding cases diagnosed between age 46 and 55 years as potentially perimenopausal. We used age at diagnosis as a proxy for menopausal status at diagnosis because menopausal status at diagnosis is difficult to determine by questionnaire, especially given the use of hormone replacement therapies; while information on age at diagnosis was available for all but 1.4% ($n = 554$) of cases, information on age at natural menopause was missing for 65.6% ($n = 26,552$) of cases of European ancestry. Similarly, although rs10235235 is a plausible candidate for association with age at menopause, we did not test this due to the limited amount of data on age at natural menopause for controls of European ancestry ($n = 11,294, 28.2\%$) and the difficulty in ascertaining whether treatment for breast cancer had influenced reported age at menopause for cases.

The strengths of our study include the large size of this combined analysis, and the availability of information on tumour characteristics for the majority of cases and on age at menarche for the majority of cases and controls. Limitations include low power of the study to examine an association between genotype and breast cancer risk for non-Europeans.

**Conclusions**

In summary, we have confirmed that rs10235235 is associated with breast cancer, have shown for the first time that rs10235235 is associated with age at menarche in controls and have suggested a potential mechanism for these associations. rs10235235, which maps to the CYP3A locus, probably tags a causal variant that affects expression of one or more CYP3A genes.

**Additional files**

- **Additional file 1:** Contains Table S1 presenting details of participating BCAC studies; Table S2 presenting rs10235235 genotypes for breast cancer cases and controls from 49 BCAC studies; Table S3 presenting availability of data on age at diagnosis, hormone receptor status, morphology, grade and nodal status for breast cancer cases from 38 European BCAC studies; Table S4 presenting availability of data on age at menarche for breast cancer cases and controls from 40 European BCAC studies; and Figure S1 showing association of the rs10235235-AG genotype with breast cancer risk for women of Asian and African-American ancestry.
- **Additional file 2:** Presents details of ethical committees that approved each study.

**Abbreviations**

- BCAC: Breast Cancer Association Consortium
- CI: confidence interval
- COGS: Collaborative Oncological Gene-environment Study
- MAF: minor allele frequency
- OR: odds ratio
- $P_{\text{trend}}$: P value for linear trend
- SNP: single nucleotide polymorphism

---

**Table 4 rs10235235 and risk of breast cancer for women of European ancestry by age at menarche**

<table>
<thead>
<tr>
<th>Age at menarche (years)</th>
<th>Cases</th>
<th>Controls</th>
<th>OR$_{\text{het}}$</th>
<th>95% CI</th>
<th>$P_1$</th>
<th>OR$_{\text{hom}}$</th>
<th>95% CI</th>
<th>$P_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤11</td>
<td>4,818</td>
<td>4,749</td>
<td>1.06</td>
<td>0.95, 1.19</td>
<td>0.30</td>
<td>1.07</td>
<td>0.67, 1.72</td>
<td>0.78</td>
</tr>
<tr>
<td>12</td>
<td>5,655</td>
<td>5,720</td>
<td>0.92</td>
<td>0.83, 1.02</td>
<td>0.10</td>
<td>0.83</td>
<td>0.54, 1.28</td>
<td>0.41</td>
</tr>
<tr>
<td>13</td>
<td>7,308</td>
<td>7,379</td>
<td>0.93</td>
<td>0.85, 1.02</td>
<td>0.11</td>
<td>0.77</td>
<td>0.54, 1.09</td>
<td>0.14</td>
</tr>
<tr>
<td>14</td>
<td>5,307</td>
<td>5,743</td>
<td>0.96</td>
<td>0.86, 1.06</td>
<td>0.42</td>
<td>0.69</td>
<td>0.45, 1.06</td>
<td>0.09</td>
</tr>
<tr>
<td>≥15</td>
<td>4,385</td>
<td>4,667</td>
<td>0.84</td>
<td>0.75, 0.94</td>
<td>0.003</td>
<td>0.81</td>
<td>0.51, 1.30</td>
<td>0.39</td>
</tr>
<tr>
<td>Total</td>
<td>27,473</td>
<td>28,258</td>
<td>0.94</td>
<td>0.90, 0.98</td>
<td>0.007</td>
<td>0.81</td>
<td>0.67, 0.98</td>
<td>0.03</td>
</tr>
</tbody>
</table>

$H_0$, null hypothesis; OR$_{\text{het}}$, odds ratio comparing rs10235235 AG genotype versus AA genotype; OR$_{\text{hom}}$, odds ratio comparing rs10235235 GG genotype versus AA genotype; $P_1$, test of $H_0$, no association between rs10235235 and breast cancer risk; $P_{\text{trend}}$, test of $H_0$, no linear trend in stratum specific estimates.
Competing interests

The authors state that they have no competing interests.

Authors' contributions

OF, FD and NO performed the statistical analyses. OF, IdSS and NJ drafted the manuscript. NJ, FD, NO, LG, MEJ, MJS, EIF, BPH, MG-C, MDo, AA, AJS, JP, IdSS and OF comprised the writing group who was responsible for the interpretation of the results and for critically reviewing the manuscript. AC, AJ, AH&W, AMa, BBlu, C-YS, DL, ES, G-C, HN, Hibre, Hlha, ILA, JC-C, J-YC, JY, LJC, LMa, BBlu, HMI, PAF, RR, SEB, TD, Mks and UH also significantly contributed to the interpretation of the results. OF, IdSS, NJ, JP, LG, DFE, MB and JW knew out the genotyping and/or data analysis. FD, NO, MEJ, MJS, EIF, BPH, JUH, MSc, GSD, CA, Mks, AB, LVM, FA, KMl, AlO, PAF, MBl, ABE, ESP, ES, IT, Mk, NM, BBlu, FMs, AS, CS, PG, TT, EC, Fme, SBE, BGN, HF, RM, MPZ, JAP, Jbne, LBa, H-C, AZ, CCD, Hibre, HMu, VA, AKD, AMe, JH, CRBl, HKs, Bla, C-J, Y-DK, The GENICA Network, HN, TAM, KA, CB, HMa, TD, NVB, NAlA, ALi, AMa, VK-VMk, JMBl, JHC, GCT, Jbne, KConfFab Investigators, Australian Ovarian Cancer Study Group, AhHw, DVBb, C-Ge, DL, DS, PH, NG, J-C-C, AR, SN, DF-J, PR, PP, BBlu, Vp, FIC, JED, XW, YS, VSP, C-GG, GS, LBlu, Cb, H, JS, MSg, FL, MDu, ST, CYF, SYC, BpC, VNk, GGA, A-LB-D, WZ, RW, KP, A-JV, MG, ILa, JAK, GG, AM, PD, JF, SIC, JlCl, MLS, PH, NS, Mho, HA, RAO, M-TeL, JUlu, AC, IBW, MWrr, SSC, WB, LBS, PDPD, AMD, MS, DK, D-YN, SkF, J-YC, MHa, HM, WYl, AT, UH, AF, TR, HUu, AJ, KJ-B, KD, Ssa, VG, PB, JM, SSI, AET, CV, DY, CYS, JCy, C-SH, M-FH, AG-N, DCT, DF, VB, CL, JD, KMa, MBl, JW, DFE, MG-C, MDo, AA and JS made substantial contributions in recruiting subjects and acquiring data, and in critically reviewing the manuscript. All authors take responsibility for the work and read and approved the final version of the manuscript.

Acknowledgements

The authors thank all of the individuals who took part in these studies and all of the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out.

ABCFS would like to thank Maggie Angelakos, Jouki Maskelli and Gilliane Dite. ABCS would like to thank Ellen van der Schoot and Sanquin Amsterdam. The ACP study wishes to thank the participants in the Thai Breast Cancer study. Special thanks also go to the Thai Ministry of Public Health (MOPH) doctors and nurses who helped with the data collection process. The study would like to thank Dr Prat Boonyawongvoj, the former Permanent Secretary of MOPH and Dr Pornthep Sarivanangaran, the Department Director-General of Disease Control who have supported the study throughout. ABCFS would like to thank Eileen Williams, Elaine Ryder-Mills and Kara Sargus. KGKFS would like to thank Susan Stock, Jacobi, Charlene. CBG would like to thank Theo Gudermann, Sven Ploog, Ralf Rothe, Ute Hamann; Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, Germany (Dr Prat Boonyawongvoj, former Permanete Secretary of MOPH and Dr Pornthep Sarivanangaran, Department Director General of Disease Control, who have supported the study throughout). KGKFS would like to thank Andrea Urban and Verena Strohmeier. ABCFS would like to thank The Swedish Medical Research Council. KGKFS would like to thank Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huissink, Annemiek Heemskerk and the Erasmus MC Family Cancer Clinic. SASSAC would like to thank The Swedish Medical Research Council. KGKFS thanks study participants and research staff for their contributions and commitment to the study. KGKFS would like to thank Meeri Otsukka and Kari Mononen. OFBCR would like to thank Teresa Selander and Nayana Weerasooriya. ORGO thanks E Kod-Warmerdam and J Blom for patient accrual, administering questionnaires and managing clinical information. The LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCODC) with the help of Dr J Molenaar. PBCS would like to thank Louise Brinton, Mark Sherman, Stephen Chanock, Neolila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatorowski, Pei Chao and Michael Stagner. pKARMa would like to thank The Swedish Medical Research Council. KGKFS would like to thank Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huissink, Annemiek Heemskerk and the Erasmus MC Family Cancer Clinic. SASSAC would like to thank The Swedish Medical Research Council. KGKFS thanks study participants and research staff for their contributions and commitment to the study. KGKFS would like to thank Sue Higham, Helen Cramp and Dan Connelly. SEARCH would like to thank The SEARCH and EPMC teams. KGKFS would like to thank the participants and research coordinator Kimberly Chua. SKKKDFSZ are grateful to all of the patients for their participation and thank the physicians and other hospital staff, scientists, research assistants and study staff who contributed to the patient recruitment, data collection and sample preparation. UKBSG thanks Breast Cancer Breathing and the Institute of Cancer Research for support and funding of the UK Breast Cancer Generations Study

Consortia members

The GENICA network: Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany (Christina Justenhoven, Hiltrud Brauch); Department of Internal Medicine, Evangelische Kliniken Bonn ggBonn, Johanniter Krankenhaus, Bonn, Germany (Yon-Dichun Ko, Christian Baisch); Institute of Pathology, University of Bonn, Germany (Hans-Peter Fischer); Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany (Ute Yamamoto); Harmonize for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, Germany (Thomas Birkenau, Beate Pesch, Sylvia Rabstein, Anne Spickeneuer); and Institute for Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany (Volker Hartl). KConfFab Investigators: David Amor, Lesley Andrews, Yoland Antill, Shane Armitage, Rosemary Balleline, Agnes Bankier, Pati Bostick, John Beilby, Barbara Bennett, Ian Bennett, Anneke Blackburn, Michael Bogwitz, Meagan Brennan, Melissa Brown, Michael Buckley, Matthew Burgess, Jo Burke, Phyllis Butow, Ian Campbell, Alice Christian, Georgia Chenevix-Trench, Christine Christine, Alison Colley, Dick Cotton, Bronwyn Culling, Margaret Cummings, Sarah-Jane Dawson, Anna DeFazio, Martin Delatycki, Rebecca Dickson, Alexander Dobrovec, Tracy Dudding, Ted Edkins, Stacey Edwards, Gelareh Farshid, Susan Fawcett, Georgina Fenton, Michael Field, James Flanagan, Peter Fong, John Forbes, Stephen Fox, Juliet French, Clara Gaff, Mac Gardner, Mike Gattas, Graham Giles, Grant Gill, Jack Goldblatt, Sian Greening, Scott Grist, Eric Haan, Marion Harris, Stewart Hart, Nick Hayward, Sue Healey, Louise Heiniger, John Hopper, Clare Hunt, Paul James, Mark Jenkins, Rick Keeford, Alexa Kidd, Belinda Kely, Judy Kir, James Kollas, Jessica Koehler, Sergio Kovalenko, Sunil Lakhani, Jennifer Leary, Geoff Lindeman, Lara Lipton, Liz Mather, Nick Milne, Geoff Molloy, Paul Monahan, Sing Mah, Susan Matthews, Nick McCallum, Peter Monahan, Andrew Morley, Alan Morris, Chris Morris, Rosemary Morison, Sara Murray, Catherine Nery, Andrew Noonan, Annette Nourski, John O’Shea, Rachel O’Shea, Andrew Parini, Frank Peckham, Rachel Petch, Sarah Vickers, Catriona Phillips, Melanie Price, Lynne Purser, Tony Reeve, Edwina Rickard, Bridget Robinson, Barney Rudzki, Elizabeth Salisbury, Christobel Saunders, Joe Sambrook, Jodi Saunon, Robin Sayer, Clare Scott, Elizabeth Scott, Rodney Scott, Adrienne Sexton, Raghu Sharma, Andrew Shelling, Peter Simpson, Melissa Southey, Amanda Spurdle, Graeme Suthers, Pamela Sykes, Jessica Taylor, Ella Thompson, Heather Thorne, Sharron Topsham, Alison Trainer, Emma Turner, Nick Veitch, Logan Walker, Paul Waring, Robin Ward, Bev Warren, Rachael Williams, Ingrid Winship, Mary Ann Young (Peter MacCallum Cancer Center, Melbourne, Australia).

The Australian Ovarian Cancer Study Group: David D Bowtell, Adele C Green, Georgia Chenevix-Trench, Anna DeFazio, Dorota Gertig, Penelope M Webb (Peter MacCallum Cancer Center, Melbourne, Australia).
Financial support

Part of this work was supported by the European Community’s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). This work was partly supported by the Canadian Institutes of Health Research for the ‘CIHR Team in Familial Risks of Breast Cancer’ program (US, DFE), and the Ministry of Economic Development, Innovation and Export Trade of Quebec – grant number PSR-SIIRI-701 (US, DFE, PH). The ABCFS and OFBCR work was supported by the United States National Cancer Institute, National Institutes of Health (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 (U11 CA64967), Northern California Cancer Center (U01 CA69417) and University of Melbourne (U11 CA69638). Samples from the NC-BCFR were processed and distributed by the Coriell Institute for Medical Research. 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 BCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. 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. JHL is a National Health and Medical Research Council (NHMRC) Australia Fellow and a Victorian Breast Cancer Research Consortium Group Leader. MCS is a NHMRC Senior Research Fellow and a Victorian Breast Cancer Research Consortium Group Leader. The ABCS study was supported by the Dutch Cancer Society (grants N01 2001-2412 and 2007-3839) and the Dutch National Genomics Initiative. The ABCS study is funded by the Breast Cancer Research Trust, UK. The work of the BBBCS was partly funded by ELAN-Fond of the University Hospital of Erlangen. BBCS is funded by Cancer Research UK and Breakthrough Breast Cancer, and acknowledges NIH funding to the NIH Biomedical Research Center and the National Cancer Research Network. BCAC is funded by CR-UK (CI287/A10118 and CI287/A1014). Meetings of the BCAC have been funded by the European Union COST programme (BM0606). DFE is a Principal Research Fellow of CR-UK. ES (BIGGS) is supported by NHMRC Comprehensive Biomedical Research Centre, Guy’s & St. Thomas’ NHS Foundation Trust in partnership with King’s College London, UK. IT is supported by the Oxford Biomedical Research Centre. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DFKZ). CGFS is supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital. CNIO-BCS was supported by the Genome Spain Foundation, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI081183 and PI08120). CTs was supported by the California Breast Cancer Act of 1993, the NIH (grants R01 CA77398 and the Lon V Smith Foundation) and the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. 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).

GC-HBRC was supported by Deutsche Krebshilfe (107054), the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Centre (DFKZ). GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/S, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DFKZ), Heidelberg, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanneskrankenhaus, Bonn, Germany. HEBCS was supported by the Academy of Finland (132473), Helsinki University Central Hospital Research Foundation, the Sigrid Juselius Foundation, the Finnish Cancer Society and the Nordic Cancer Union. HERPACC was supported by a Grant-in-Aid for Scientific Research on Priority Areas and on Innovative Area from the Ministry of Education, Science, Sports, Culture and Technology of Japan and by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from Ministry Health, Labour and Welfare of Japan. HMBCS was supported by short-term fellowships from the German Academic Exchange Program (NWB) and the Friends of Hannover Medical School (NWB). KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the Academy of Finland and by the strategic funding of the University of Eastern Finland. McKonFab is supported in part by funding from the National Breast Cancer Foundation, the NHMRC, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia and the Cancer Foundation of Western Australia. The McKonFab Clinical Follow Up Study was funded by the NHMRC (145684, 288704, 454508). Financial support for the AOCs was provided by the United States Army Medical Research and Materiel Command (DAMD17-01-1-0729), the Cancer Council of Tasmania and Cancer Foundation of Western Australia and the NHMRC (199600). GC-T and P Webb are supported by the NHMRC. LAABC is supported by grants (1R8-0287, 3PR-0102, 5PB-0018, 10PB-0008) from the California Breast Cancer Research Program. Incident breast cancer cases were collected by the US Cancer Surveillance Program (CSP), which is supported under subcontract by the California Department of Health. CSP is also part of the National Cancer Institute’s Division of Cancer Prevention and Control Surveillance, Epidemiology, and End Results Program, under contract number NO1CN25403. LMBC is supported by the ’Stichting tegen Kanker’ (232-2008 and 196-2010). DL is supported by the KULPEV/10/016-SymBoSyll.

The MARIE study was supported by the Deutsche Krebshilfe e.V. (70-2892-BR-I), the Hamburg Cancer Society, the German Cancer Research Center and the genotype work in part by the Federal Ministry of Education and Research (BMBF) Germany (01H0402). MBCSG was funded by grants from Italian Association for Cancer Research (AIRC, IG 8713), and by Italian citizens who allocated the 5 x 1000 share of their tax payment in support of the Fondazione I.R.C.C. Istituto Nazionale dei Tumori, according to Italian laws (INT-Institutional strategic projects 5x1000). MCBSCS was supported by the NIH grants CA116167 and CA128978, an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), the Breast Cancer Research Foundation, and a generous gift from the David F and Margaret T Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation of Western Australia. Collection of the MARIE dataset was funded by VicHealth and Cancer Council Victoria. MCCS was further supported by Australian NHMRC grants 200057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. MEC was supported by NIH grants CA53644, CA54281, CA09758 and CA132839. The work of MLTGBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the ‘CIHR Team in Familial Risks of Breast Cancer’ program (grant number CRN-87521) and the Ministry of Economic Development, Innovation and Export Trade (grant number PSR-SIIRI-701). MYBRCA is funded by grants from the Malaysian Ministry of Science, Technology and Innovation (MOSTI), Malaysian Ministry of Higher Education (UM/CIHR/MOHE/06) and Cancer Research Initiatives Foundation (CARIF). Additional controls were recruited by the Singapore Eye Research Institute, which was supported by a grant from the Biomedical Research Council, Singapore (BMRC/08/1/35/19550) and the National Medical Research Council, Singapore (NMRC/CG/SGER/2010). NBCS was supported by grants from the Norwegian Research council, 155218/40, 175240/510 to A-LB-D, FUGE-NFR 181600/11 to VNK and a Swiss Bridge Award to A-LB. NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA64885. OBCS was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland, the University of Oulu, and the Oulu University Hospital. The ORGO study was supported by the Dutch Cancer Society (RUL. 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The pAKARMA study was supported by Märt and Hans Rausings Initiative Against Breast Cancer. RBCS was funded by the Dutch Cancer Society (DDH 2004-3124, DDH 2009-4318). The NASABC study was funded by funding from the Cancer Technology and Research of Singapore (A*STAR), the US NIH and the Susan G Komen Breast Cancer Foundation. SBCCS was supported primarily by NIH grants R01CA64277, R01CA148667, and R37CA70867. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA64885. SBCS was supported by Yorkshire Cancer Research S295, S299, S305PA. SCSS is supported by a grant from the
National Institutes of Health (R01 CA092447). Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry, South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; and Arkansas Department of Health, Cancer Registry, Little Rock. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry, which participates in the National Program of Cancer Registries of the CDC. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry. SEARCH is funded by programme grants from Cancer Research UK (C490/A10124 and C8197/A10123) and NIH grant 5U10CA082617-06. SEBCS was supported by the Korea Health 21 R&D Project (A033001), Ministry of Health and Welfare, Republic of Korea. SGBCC is funded by the National Medical Research Council of Thailand. TBCS was supported by The National Cancer Institute, Thailand. TNBCS was supported by an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), the Breast Cancer Research Foundation, a generous gift from the David F and Margaret T Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation; The Stefanie Spielman Breast Cancer Fund and the OSI Comprehensive Cancer Center; the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF) – Research Funding Program of the General Secretariat for Research & Technology (ARISTEIA). TBCIS is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences, Academia Sinica, Taiwan. UKBGS is funded by Breakthrough Breast Cancer and the Institute of Cancer Research (ICR). ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.

Author details
1Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW6 6JR, UK. 2Division of Breast Cancer Research, The Institute of Cancer Research, 237 Fulham Road, London SW6 6LB, UK. 3Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. 4Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Way, Belmont, Sutton, Surrey SM2 5NG, UK. 5The Academic Unit of Biochemistry, The Royal Marsden Hospital, Fulham Road, London SW3 6JJ, UK. 6Centre for Molecular, Environmental, Genetic and Cancer Epidemiology, University of Melbourne, 1-100 Gratton Street, Parkville, Melbourne, Victoria 3010, Australia. 7Genetic Epidemiology Department, Department of Pathology, The University of Melbourne, 1-100 Gratton Street, Parkville, Melbourne, Victoria 3010, Australia. 8Division of Molecular Pathology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Pneumonie Inst. 121, 1066 CX Amsterdam, The Netherlands. 9SamfunN grant and Centre Grant (NMRCA/CG/CNICS/2010). Additional controls were recruited by the Singapore Consortium of Cohort Studies-Multi-ethnic cohort (SCCS-MEC), which was funded by the Biomedical Research Council, grant number: 05/1/21/19/425. SKOMAF is supported by the DKFZ, Heidelberg, Germany. KB-B (SZBCS) is a fellow of International PhD program, Postgraduate School of Molecular Medicine, Warsaw Medical University, supported under the Polish Foundation of Science. TBCS was funded by The National Cancer Institute, Thailand. TNBCS was supported by an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), the Breast Cancer Research Foundation, a generous gift from the David F and Margaret T Grohne Family Foundation and the Ting Tsung and Wei Fong Chao Foundation; The Stefanie Spielman Breast Cancer Fund and the OSI Comprehensive Cancer Center; the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF) – Research Funding Program of the General Secretariat for Research & Technology (ARISTEIA). TBCIS is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences, Academia Sinica, Taiwan. UKBGS is funded by Breakthrough Breast Cancer and the Institute of Cancer Research (ICR). ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.
References


