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

Johnson, Nichola; Dudbridge, Frank; Orr, Nick; Gibson, Lorna; Jones, Michael E.; Schoemaker, Minouk J.; Folkerd, Elizabeth J.; Haynes, Ben P.; Hopper, John L.; Southey, Melissa C.; Dite, Gillian S.; Apicella, Carmel; Schmidt, Marjanka K.; Broeks, Annegien; Van't Veer, Laura J.; Atsma, Femke; Muir, Kenneth; Lophatananon, Artitaya; Fasching, Peter A.; Beckmann, Matthias W.; Ekici, Arif B.; Renner, Stefan P.; Sawyer, Elinor; Tomlinson, Ian; Kerin, Michael; Miller, Nicola; Burwinkel, Barbara; Marme, Frederik; Schneeweiss, Andreas; Sohn, Christof; Guenel, Pascal; Truong, Therese; Cordina, Emilie; Menegaux, Florence; Bojesen, Stig E.; Nordestgaard, Borge G.; Flyger, Henrik; Milne, Roger; Zamora, M. Pilar; Aras Perez, Jose Ignacio; Benitez, Javier; Bernstein, Leslie; Anton-Culver, Hoda; Ziegas, Argyrios; Dur, Christina Clarke; Brenner, Hermann; Mueller, Heiko; Arndt, Volker; Dieffenbach, Aida Karina; Meindl, Alfons

Published in:
Breast Cancer Research

DOI:
10.1186/bcr3662

Published: 2014-01-01

Link to publication

Citation for published version (APA):
Genetic variation at CYP3A is associated with age at menarche and breast cancer risk: a case-control study


*Correspondence: nichola.johnson@icr.ac.uk
†Equal contributors
1Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK
2Division of Breast Cancer Research, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK
Full list of author information is available at the end of the article

© 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.
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 (\(P_{\text{trend}} = 0.02\)). There was no evidence of effect modification by tumor characteristics. rs10235235 was, however, associated with age at menarche in controls (\(P_{\text{trend}} = 0.005\)) but not cases (\(P_{\text{trend}} = 0.97\)). Consequently the association between rs10235235 and breast cancer risk differed according to age at menarche (\(P_{\text{het}} = 0.02\)): 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_{nom} = 0.81, 95% CI 0.51, 1.30; \(P_{\text{trend}} = 0.002\)) but not for those who had their menarche age ≤11 years (OR_{het} = 1.06, 95% CI 0.95, 1.19, OR_{nom} = 1.07, 95% CI 0.67, 1.72; \(P_{\text{trend}} = 0.29\)).

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.
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 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 discordant 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 (MYBRCRA; 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 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 likelihood ratio test. Heterogeneity in ORs between studies within each subgroup (European, Asian and African-American), and between subgroups, was assessed using the Cochran Q statistic and quantified using the I² measure [20].

Analyses stratified by oestrogen receptor status (+/−), progesterone receptor status (+/−), morphology (ductal or lobular), grade (1,2,3), lymph node involvement (+/−) or age at diagnosis (≤50 and >50 years) were restricted to studies of European ancestry due to the small number of studies of Asian and African-American ancestry. In addition, studies were excluded if they had selected cases on the basis of the stratifying variable, or had collected data on that variable for less than 5% of cases or less than 10 cases in total. Availability of data for each of the stratifying variables in each study is shown in Table S3 in Additional file 1. To assess the relationship between each of the stratifying variables and genotype, stratum-specific ORs were calculated using logistic regression. Cases in each stratum were compared with all control subjects, adjusted for study and principal components. Case-only logistic regression was used to test for heterogeneity between strata (binary stratifying variables) or across strata (stratifying variables with three or more strata). P values were estimated using likelihood ratio tests with one degree of freedom.

We assessed whether rs10235235 was associated with age at menarche in cases and controls separately. Studies that had not collected data on age at menarche in both cases and controls were excluded (Table S4 in Additional file 1). We used linear regression, adjusted for principal components and study, to estimate the relationship between age at menarche (years) and rs10235235 genotype (0, 1, 2 rare alleles) and logistic regression adjusted for principal components and study to estimate the 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 (Phet) = 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 (Phet = 1 × 10⁻²²) that was accounted for by the three Finnish studies (HEBCS, MAF = 0.15; KBCP, MAF = 0.21; and OBCS, MAF = 0.15; Phet = 0.01); no evidence of heterogeneity remained after taking account of these studies (MAF = 0.087 (95% CI = 0.085, 0.089); Phet = 0.23).

Relative to Europeans, the overall MAF was higher for African-Americans (0.213, 95% CI = 0.195, 0.232; Phet = 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 (Phet = 4 × 10⁻¹⁴).

We assessed whether rs10235235 was associated with age at menarche in cases and controls separately. Studies that had not collected data on age at menarche in both cases and controls were excluded (Table S4 in Additional file 1). We used linear regression, adjusted for principal components and study, to estimate the relationship between age at menarche (years) and rs10235235 genotype (0, 1, 2 rare alleles) and logistic regression adjusted for principal components and study to estimate the 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 (Phet) = 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 (Phet = 1 × 10⁻²²) that was accounted for by the three Finnish studies (HEBCS, MAF = 0.15; KBCP, MAF = 0.21; and OBCS, MAF = 0.15; Phet = 0.01); no evidence of heterogeneity remained after taking account of these studies (MAF = 0.087 (95% CI = 0.085, 0.089); Phet = 0.23).

Relative to Europeans, the overall MAF was higher for African-Americans (0.213, 95% CI = 0.195, 0.232; Phet = 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 (Phet = 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 (Ptrend) = 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 (Phet = 0.44, Pt = 1.9% and Phet = 0.76, I² = 0.0% for heterozygote and homozygote OR estimates respectively). There was, however, some evidence of effect modification by morphology for either OR estimate (Phet = 0.44, Pt = 1.9% and Phet = 0.76, I² = 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 1,114 cases and no carriers among 1,219 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 (Phet = 0.51).

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

The SNP rs10235235 maps to a locus (CYP3A) 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 RPCI) could not be estimated because there were no GG homozygotes among cases or 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 (\(P_{\text{het}} = 0.89\); Table 2), and excluding cases who were diagnosed between age 46 and 55 as potentially perimenopausal did not alter this result (\(P_{\text{het}} = 0.28\)). 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 (\(P_{\text{het}} = 0.02\); Table 4); being a carrier of one or two rare alleles of rs10235235 was associated with an estimated 16% (OR\(_{\text{het}} = 0.84\), 95% CI = 0.75, 0.94; \(P = 0.003\)) or 19% (OR\(_{\text{hom}} = 0.81\), 95% CI = 0.51, 1.30; \(P = 0.39\)) (\(P_{\text{trend}} = 0.002\)) reduction in breast cancer risk for women who had their menarche at ages ≥15 years but there was no evidence of reduction for those with a menarche at age ≤11 years (OR\(_{\text{het}} = 1.06\), 95% CI = 0.95, 1.19; \(P = 0.30\) and OR\(_{\text{hom}} = 1.07\), 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 (n)</th>
<th>Controls (n)</th>
<th>OR(_{\text{het}})</th>
<th>95% CI</th>
<th>(P_{\text{het}})</th>
<th>OR(_{\text{hom}})</th>
<th>95% CI</th>
<th>(P_{\text{hom}})</th>
<th>(P_{\text{trend}})</th>
<th>(P_{\text{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>
<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>
<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>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>38,970(^a)</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>0.83</td>
<td></td>
</tr>
<tr>
<td>PR status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<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>
<td></td>
<td></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>
<td></td>
<td></td>
</tr>
<tr>
<td>NK</td>
<td>12,111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>38,801(^b)</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>
<td>0.19</td>
<td></td>
</tr>
</tbody>
</table>

**Morphology**

| Ductal            | 22,123    | 31,803       | 0.98                 | 0.93, 1.02| 0.30                | 0.74                 | 0.61, 0.90| 0.003               |                     |                     |
| Lobular           | 3,921     | 31,803       | 1.07                 | 0.98, 1.17| 0.14                | 0.91                 | 0.64, 1.27| 0.57                |                     |                     |
| Other and NK      | 5,995     |              |                      |          |                     |                      |          |                     |                     |                     |
| Total             | 32,039    | 31,803       | 0.99                 | 0.95, 1.04| 0.64                | 0.77                 | 0.64, 0.92| 0.004               | 0.03                |                     |

**Grade**

| Grade 1           | 5,944     | 37,285       | 0.97                 | 0.90, 1.05| 0.46                | 0.86                 | 0.65, 1.15| 0.31                |                     |                     |
| Grade 2           | 13,427    | 37,285       | 1.00                 | 0.95, 1.06| 0.92                | 0.80                 | 0.63, 0.98| 0.04                |                     |                     |
| Grade 3           | 8,638     | 37,285       | 0.98                 | 0.92, 1.05| 0.58                | 0.61                 | 0.46, 0.82| 0.001               |                     |                     |
| NK                | 8,769     |              |                      |          |                     |                      |          |                     |                     |                     |
| Total             | 36,778    | 37,285       | 0.99                 | 0.95, 1.03| 0.56                | 0.76                 | 0.64, 0.90| 0.001               | 0.63                |                     |

**Nodal status**

| Node-negative     | 17,463    | 37,836       | 0.98                 | 0.93, 1.03| 0.47                | 0.86                 | 0.71, 1.04| 0.12                |                     |                     |
| Node-positive     | 10,746    | 37,836       | 0.98                 | 0.92, 1.04| 0.46                | 0.72                 | 0.57, 0.93| 0.01                |                     |                     |
| NK                | 9,359     |              |                      |          |                     |                      |          |                     |                     |                     |
| Total             | 37,568    | 37,836       | 0.98                 | 0.94, 1.02| 0.31                | 0.81                 | 0.68, 0.96| 0.02                | 0.51                |                     |

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\(_{\text{het}}\), odds ratio comparing rs10235235 AG genotype versus AA genotype; H\(_0\), null hypothesis; NK, not known; OR\(_{\text{hom}}\), odds ratio comparing rs10235235 GG genotype versus AA genotype; \(P_{\text{trend}}\), test of H\(_0\) no association between rs10235235 and breast cancer risk; \(P_{\text{het}}\), test of H\(_0\) no difference between stratum specific estimates for variables with two strata or test of H\(_0\) no linear trend in stratum specific estimates for variables with three strata. \(^a\)Excludes seven studies that selected all ER-negative cases (CTS, DEMOKRITOS, NBCS, NBHS, OSU, RPCI and SKKD/KFZS) and one study (PBCS) that selected all ER-positive cases. \(^b\)Excludes seven studies that selected all PR-negative cases (CTS, DEMOKRITOS, NBCS, NBHS, OSU, RPCI and SKKD/KFZS).
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>P1</th>
<th>ORhom 95% CI</th>
<th>P1</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>0.68</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>

Table 3 Association of rs10235235 with age at menarche for women of European ancestry by case-control status

<table>
<thead>
<tr>
<th>rs10235235 genotype</th>
<th>Cases</th>
<th>Age at menarche (years)</th>
<th>P_trend</th>
<th>Controls</th>
<th>Age at menarche (years)</th>
<th>P_trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>22,954</td>
<td>12.83</td>
<td>23,383</td>
<td>12.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>4,312</td>
<td>12.83</td>
<td>4,627</td>
<td>13.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>207</td>
<td>12.83</td>
<td>248</td>
<td>13.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27,473</td>
<td>12.83</td>
<td>28,258</td>
<td>12.96</td>
<td>0.005</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

This study of more than 47,000 breast cancer cases and 47,000 controls has confirmed that rs10235235, mapping to 7q22.1 (CYP3A), 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 CYP3A 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 CYP3A 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, r^2 < 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 ($P_{het} = 0.03$). 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; $P_{trend}$: 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$_{het}$ 95% CI</th>
<th>$P_1$</th>
<th>OR$_{hom}$ 95% CI</th>
<th>$P_1$</th>
<th>$P_{het}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤11</td>
<td>4,818</td>
<td>4,749</td>
<td>1.06 0.95, 1.19</td>
<td>0.30</td>
<td>1.07 0.67, 1.72</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>5,655</td>
<td>5,720</td>
<td>0.92 0.83, 1.02</td>
<td>0.10</td>
<td>0.83 0.54, 1.28</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>7,308</td>
<td>7,379</td>
<td>0.93 0.85, 1.02</td>
<td>0.11</td>
<td>0.77 0.54, 1.09</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5,307</td>
<td>5,743</td>
<td>0.96 0.86, 1.06</td>
<td>0.42</td>
<td>0.69 0.45, 1.06</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>≥15</td>
<td>4,385</td>
<td>4,667</td>
<td>0.84 0.75, 0.94</td>
<td>0.003</td>
<td>0.81 0.51, 1.30</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27,473</td>
<td>28,258</td>
<td>0.94 0.90, 0.98</td>
<td>0.007</td>
<td>0.81 0.67, 0.98</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

$H_0$, null hypothesis; OR$_{het}$, odds ratio comparing rs10235235 AG genotype versus AA genotype; OR$_{hom}$, odds ratio comparing rs10235235 GG genotype versus AA genotype; $P_1$, test of $H_0$ no association between rs10235235 and breast cancer risk; $P_{het}$, test of $H_0$ no linear trend in stratum specific effects.
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, ME, MJS, EIF, BPH, MG-C, MD, AA, AJS, JP, IdSS and OF comprised the writing group that was responsible for the interpretation of the results and for critically reviewing the manuscript. AC, AJ, AHW, ABu, C-YS, DL, ES, GC-T, HN, HBre, HMa, ILA, JCC, J-YC, JLH, LBS, MBK, HMI, PAF, PF, RB, SEB, TD, MKS and UH also significantly contributed to the interpretation of the results. OF, IdSS, NJ, LP, JG, DFE, MBM and JW conceived of the original design of the study and participated in subject recruitment and in acquisition of data. JBen, AG-N, RM, DCT, DF, VB, CL, JD, Js and KMU carried out the genotyping and/or data analysis. FD, NO, MEI, MJS, EIF, BPH, JUH, MCS, GSD, CA, MKS, AB, LVM, FA, KML, AL, PAF, MWB, ABE, SPB, ES, IT, MK, NM, BBlu, FMa, AS, CS, PG, TT, EC, Fme, SEB, BGF, HM, RM, MP, JAP, JLI, Bde, Ha-C, AZ, CCD, HBl, HMa, VA, AKD, Ame, CHJ, CRB, HKa, BJa, C-J, YDK, The GENICA Network, HN, TAM, KA, CB, HMa, TD, NVB, NNA, ALI, AMa, VK, V-MK, JMH, GCT, JBl, KConfab Investigators, Australian Ovarian Cancer Study Group, AHM, DVB, C-Gr, DL, DS, PN, HN, JCC-C, AR, SN, DF-J, PR, PP, BBp, VP, FIC, JED, JXW, JZ, VSP, CQG, GS, LBA, ChC, JS, MSG, FL, MDU, ST, CHY, SYC, YP, BCK, VNN, GGA, A-LB-D, WZ, RW, KP, AJ-V, MG, ILA, JAK, GG, AMD, PF, J, SJL, JIS, MES, PH, NS, MHo, RA, HAO, MTL-JLI, AC, IBW, MVWR, SSC, WB, LBS, PDPD, AMD, MS, DK, D-YN, SKP, J-YC, MHa, HM, WYH, AT, UH, AF, TR, HUJ, AJ, JL, K-B, KD, SSa, VG, PB, JM, SSJ, AET, CV, D, CY, CYS, J-CH, S-F, M-FH, AG-N, DCT, DF, VB, CL, JD, KM, MBK, JW, DFE, MG-C, MD, AA and AUU 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, Judi Maskell and Gillian 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. ABFS would like to thank Maggie Angelakos, Judi Maskell and Gillian Dite. The Authors state that they have no competing interests.
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 (U5, DFE), and the Ministry of Economic Development, Innovation and Export Trade of Quebec – grant number P8R-SIRI-701 (U5, 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 (U01 CA69467), Northern California Cancer Center (U01 CA69417) and University of Melbourne (U01 CA69638). Samples from the NC-BFCR were processed and distributed by the Correll 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. J LH 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 NH 2001-24123 and 2007-3089) and the Dutch National Genomics Initiative. The ACP study is funded by the Breast Cancer Research Trust, UK. The work of the BBCS was partly funded by ELAN-Fond of the University Hospital of Erlangen. BCCS is funded by Cancer Research UK and Breakthrough Breast Cancer, and acknowledges NHF funding to the NIH Biomedical Research Center and the National Cancer Research Network. BCAC is funded by CR-UK (C1287/A10118 and C1287/A10214). 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 NHF 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 was supported by the Chief Physician Johan Bosenup 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 (PI081583 and PI081120). CTs was supported by the California Breast Cancer Act of 1993, the NIH (grants RO1 CA77398 and the Lon V Smith Foundation (LVS39420)) 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 VERRDI study, which was supported by a grant from the German Cancer Foundation (Deutsche Krebshilfe).

GC-HBRC was supported by Deutsche Krebshilfe (107054), the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DFKZ). GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany (01KWK075/S, 01KWK976/8, 01KWK977/0 and 01KWK0114, 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, Johanniter Krankenhaus, Bonn, Germany). HERBS 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 funded 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 (DAAD) and the Friends of Hannover Medical School (NVh). 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. kConFab is supported 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 kConFab 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 Council of Western Australia and the NHMRC (199600). GC-T and P Webb are supported by the NHMRC. LAABC is supported by grants (18R-0287, 3PB-0102, 5PB-0018, 10PB-0008) from the California Breast Cancer Research Program. Incident breast cancer cases were collected by the USC 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 N01CN25403. LMBC is supported by the 'Stichting tegen Kanker' (232-2008 and 196-2010). DL is supported by the KULPEV/10/016-SymBoSys.

The MARe 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 (01KWK0402), MBCSG was funded by grants from Italian Association for Cancer Research (AIRC, IG 8213) and by Italian citizens who allocated the 5 x 1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale dei Tumori, according to Italian laws (INT-Institutional strategic projects 5x1000). MBCBS was supported by the NIH grants CA116167 and CA28978, 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. The kConFab cohort recruitment 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 CA63464, CA54281, CA09783 and CA132839. The work of MtlGBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the ‘Clínica de Familia’ Project of Breast Cancer program (grant number CRN-87251) and the Ministry of Economic Development and Innovation and Export Trade (grant number P8R-SIRI-701). MYBCRA is funded by grants from the Malaysian Ministry of Science, Technology and Innovation (MOSTI), Malaysian Ministry of Higher Education (UMCA/H/CRMOHE/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 (BMRC81/135/19/550) and the National Medical Research Council, Singapore (NMRC/CG/SERI/2010). NBCS was supported by grants from the Norwegian Research council, 155218/ VA0, 175240/S10 to A-LB-D, FUGE-NFR 181600/V11 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 CA68485. OBBCS 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 pKARRMA study was supported by Mait and Hans Rausing Initiative Against Breast Cancer. RBCS was funded by the Dutch Cancer Society (DDH4 2004-3124, DDH4 2009-4318). The pKARRMA study was funded by the Agency for Science, Technology and Research of Singapore (A*STAR), the US NIH and the Susan G Komen Breast Cancer Foundation. SBCG 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 CA68485. SBCS was supported by Yorkshire Cancer Research S295; S299; S305PA. SCCS 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 (C490/1A10124 and C8197/ A10123) and NIHR grant S0U1CA082617-16. 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 Singapore (SIMRC) through the Singapore Consortium of Cohort Studies-Multi-ethnic cohort (SCCS-MEC), which was funded by the Biomedical Research Council, grant number: 05/121/19/425. SKIODFZ2 is supported by the DFKZ, Heidelberg, Germany. K-B (SBZCS) is a fellow of International PhD program, Postgraduate School of Molecular Medicine, Warsaw Medical University, supported by the Polish Foundation of Science. TBCS was funded by The National Cancer Institute, Thailand. TNBCBS 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 OSU Comprehensive Cancer Center; the European Union (European Social Fund – ESF) and Greek national funds through the operational Programme Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) – Research Funding Program of the General Secretariat for Research & Technology: ARISTEIA. TWBCS is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences, Academia Sinica, Taiwan. UMBS 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 SW3 6JB, UK. 2Division of Breast Cancer Research, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, 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 Road, Belmont, Sutton, Surrey SM2 5NG, UK. 5The Academic Department of Biochemistry, The Royal Marsden Hospital, Fulham Road, London SW3 6JJ, UK. 6Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, 1-100 Grattan Street, Parkville, Melbourne, Victoria 3010, Australia. 7Genetic Epidemiology Department, Department of Pathology, The University of Melbourne, 1-100 Grattan Street, Parkville, Melbourne, Victoria 3010, Australia. 8Division of Molecular Pathology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, P.O. Box 21, 1009 AZ Amsterdam, The Netherlands. 9Samman, Radboud Universiteit Nijmegen, 6525 GA, Nijmegen, The Netherlands. 10Warwick Medical School, University of Warwick, Coventry CV4 7AJ, UK. 11University Breast Center, Department of Gynecology and Obstetrics, University Hospital Erlangen, Postfach 2306, D-91012 Erlangen, Germany. 12David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, University of California, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA. 13Institute of Human Genetics, Friedrich Alexander University Erlangen-Nuremberg, Schloßplatz 4, 91054 Erlangen, Germany. 14Division of Cancer Studies, NIHR Comprehensive Biomedical Research Centre, Guy’s & St. Thomas’ NHS Foundation Trust in partnership with King’s College London, Guy’s Hospital, Great Maze Pond, London SE1 9RT, UK. 15Welcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. 16Oxford Biomedical Research Centre, University of Oxford, The Churchill Hospital, Old Road, Headington OX3 7LE Oxford UK. 17Surgery, Clinical Science Institute, Galway University Hospital and National University of Ireland, University Road, Galway, Ireland. 18Department of Obstetrics and Gynecology, University of Heidelberg, Vossstrasse 9, 69115 Heidelberg, Germany. 19Unit Molecular Epidemiology CBO80, German Cancer Research Center, DFKZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 20National Centre for Tumor Diseases, University of Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany. 21Inserm (National Institute of Health and Medical Research), CESP (Centre for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, 101 rue de Tolbiac, Villejuif, 75654 Paris, France. 22University Paris-Sud, UMR1018, 101 rue de Tolbiac, Villejuif, 75654 Paris, France. 23Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, 2730 Herlev, Copenhagen, Denmark. 24Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, 2730 Herlev, Copenhagen, Denmark. 25Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, 2730 Herlev, Copenhagen, Denmark. 26Genetic and Molecular Epidemiology Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Calle de Melchor Fernandez Almagro, 3, 28029 Madrid, Spain. 27Centre for Investigation en Red de Enfermedades Raras (CIERRE), Calle de Melchor Fernandez Almagro, 3, 28029 Madrid, Spain. 28Division of Cancer Etiology, Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA, USA. 29Department of Epidemiology, School of Medicine, 224 Irvine Hall, University of California Irvine, Irvine, California 92697-7550, USA. 30Cancer Prevention Institute of California, 2201 Walnut Avenue, Suite 300, Fremont, California 95508, USA. 31Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Im Neuenheimer Feld 280, 69121 Heidelberg, Germany. 32German Cancer Consortium (DKTK), Im Neuenheimer Feld 280, 69121 Heidelberg, Germany. 33Clinic of Gynecology and Obstetrics, Division of Tumor Genetics, Klinikum rechts der Isar, Technical University Munich, Ismaninger Strasse 22, D-81675 Munich, Germany. 34Institute of Human Genetics, University of Heidelberg, Im Neuenheimer Feld 366, 69121 Heidelberg, Germany. 35Division of Molecular Gyneco-Oncology, Department of Gynaecology and Obstetrics, Center of Molecular Medicine Cologne (CMMMC), University Hospital of Cologne, ZMK-Forschungsgebäude, Robert-Koch-Strasse 21, 50931 Cologne, Germany. 36Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Robert Bosch Stiftung GmbH, Heidelheiferstrasse 31, 70184 Stuttgart, Germany. 37University of Tübingen, Geschwister-Scholl-Platz, 72074 Tübingen, Germany. 38Department of Internal Medicine, Evangelische Kliniken Bonn GGmbH, Jochaniter Krankenhaus, 53113 Bonn, Germany. 39Department of Obstetrics and Gynecology, Helsinki University Central Hospital, University of Helsinki, P.O. Box 1401, amountsaniku, FIN-00029 Helsinki, Finland. 40Department of Clinical Genetics, Helsinki University Central Hospital, P.O. Box 1401, amountsaniku, FIN-00029 Helsinki, Finland. 41Department of Oncology, Helsinki University Central Hospital, P.O. Box 1401, amountsaniku, FIN-00029 Helsinki, Finland. 42Department of Epidemiology and Prevention, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan. 43Department of Obstetrics and Gynaecology, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. 44N.N. Alexandrov Research Institute of Oncology and Medical Radiology, 223040, p. Lesnny, Minsk, Belarus. 45Department of Molecular Medicine and Surgery, Karolinska Institutet, Solnavägen 1, 171 77 Stockholm, Sweden. 46Department of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Yliopistonranta 1, P.O. Box 1627, FI-70211 Kuopio, Finland. 47Biocenter Kuopio, Cancer Center of Eastern Finland, University of Eastern Finland, Yliopistonranta 1, P.O. Box 1627, FI-70211 Kuopio, Finland. 48Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, P.O. Box 100, FI-70029 Kuopio, Finland. 49Department of Genetic, Queensland Institute of Medical Research, 300 Herston Rd, Herston, Brisbane Queensland 4006, Australia. 50Department of Preventive Medicine, Keck School of Medicine,
References


Submit your next manuscript to BioMed Central and take full advantage of:

• Convenient online submission
• Thorough peer review
• No space constraints or color figure charges
• Immediate publication on acceptance
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit