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

Nichola Johnson1,2*, Frank Dudbridge3, Nick Orr1,2, Lorna Gibson3, Michael E Jones4, Minouk J Schoemaker4, Elizabeth J Folkerd5, Ben P Haynes5, John L Hopper6, Melissa C Southey7, Gillian S Dite6, Carmel Apicella6, Marjanka Schmit6, Annegien Broeks3, Laura J Van’t Veer9, Femke Atsma9, Kenneth Mui10, Aratiya Lophatananon10, Peter A Fasching11,12, Matthias W Beckmann11, Arif B Ekici13, Stefan P Renner11, Elin Sawyer14, Ian Tomlinson15,16, Michael Kerin17, Nicola Miller17, Barbara Burwinkel18,19, Frederik Marme18, Andreas Schneeweiss18, Christof Sohn18,20, Pascal Guéner21,22, Therese Truong21,22, Emilie Cordina21,22, Florence Menegaux21,22, Stig E Bojesen23,24, Børge G Nordestgaard23,24, Henrik Flyger25, Roger Milne26, M Pilar Zamora27, Jose Ignacio Arias Perez28, Javier Benitez29,30, Leslie Bernstein31, Hoda Anton-Culver32, Argyrios Zogas32, Christina Clarke Dur33, Hermann Brenner34,35, Heiko Müller34, Volker Arndt34, Aida Karina Dieffenbach34,35, Alfons Meindl33, Joerg Heil18, Claus R Bartram37, Rita K Schmutzle38, Hilbrud Brauch39,40, Christina Justenhoven39,40, Yon-Dschun Ko41, The GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network, Heli Nevanlinna42, Taru A Muranen42, Kristinna Aittomäki43, Carl Blomqvist44, Keitaro Matsuo45, Thilo Dörk46, Natalia V Bogdanova46, Natalia N Antonenkova47, Annika Lindblom49, Arto Mannervirta50,52, Vesa Kataja50,51,53, Veli-Matti Kosma50,51,52, Jaana M Hartikainen50,51,52, Georgia Chenevix-Trench,54, Jonathan Beesley54, kConFab Investigators, Australian Ovarian Cancer Study Group, Anna H Wu55, David Van den Berg55, Chiu-Chen Tseng55, Diether Lambrechts56,57, Dominiek Smets65,67, Patrick Neven58, Hans Wildiers58, Jenny Chang-Claude59, Anja Rudolph59, Stefan Nickels59, Dieter Flesch-Jayns60,61, Paolo Radice52, Paolo Peterlongo62,63, Bernardo Bonanni54, Valeria Pensotti63,65, Fergus J Couch66, Janet E Olson67, Xianshu Wang66, Zachary Fredericksen67, Vernon S Pankratz67, Graham G Giles6,68, Gianluca Severi6,68, Laura Baglietto6,68, Chris Haiman6, Jacques Simard6, Mark S Goldber68, France Labrèche70, Martine Dumont71, Penny Soucy71, Soo Teo72,73, Cheng Har Yip72, Sze Yee Phuah72,73, Belinda K Corones74, Vessela N Kristensen75,76, Grethe Grenaker Alnæs76, Anne-Lise Børresen-Dale75,76, Wei Zheng77, Robert Winquist78, Katri Pylkas78, Arja Jukkola-Vuorinen79, Mervi Grip80, Irene L Andrus61,82, Julia A Knight61,83,84, Gerd Glendening80, Anna Marie Mulligan85,86, Peter Devilee87,88, Stephen John Chanoir88, Jolanta Lissowska89, Mark E Sherman88, Per Hall90, Nils Schook90, Maartje Hooring91, Antoinette Hollestelle92, Rogier A Oldenburg93, Madeleine Tilanus-Linthorst93, Jianjun Liu94, Angie Cox95, Ian W Brock95, Malcolm WR Reed95, Simon S Cross97, William Blot79,80, Lisa B Signorello96,98,100,101, Paul DP Pharoah102, Alison M Dunning102, Mitul Shah103, Daeehe Kang103, Dong-Young Noh103, Sue K Park104,105,106, Ji-Youl Choi103, Mikael Hartman107,108,109, Hui Miao100,109, Wei Yen Lim100,109, Anthony Tang110, Ute Hamann111, Asta Forst112,113, Thomas Rüdiger114, Hans Ulrich Ulmer115, Anna Jakubowska116, Jan Lubinski116, Katarzyna Jaworska-Bieniek116,117, Katarzyna Durda116, Suleeporn Sangrajrang118, Valerie Gaborieau119, Paul Brennan119, James McKay119, Susan Slager67, Amanda E Toland120, Celine Vachon67, Drakoulis Yannoukakos121,122

* 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.
Chen-Yang Shen1,2,13, Jyh-Cheng Yu14, Chiu-Sheng Huang15, Ming-Feng Hou16,17, Anna González-Neira19, Daniel C Tessier12,6, Daniel Vincent12,8, Francois Bacot12,8, Craig Luccarini102, Joe Dennis129, Kyriaki Michailidou129, Manjeet K Bolla129, Jean Wang129, Douglas F Easton102,129,Montserrat García-Closas1,2,4, Mitch Dowsett5,6, Alan Ashworth1,2, Anthony J Swerdlow12,4, 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 0.95, 1.19, 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 CYP3A5, was associated with a reduction of 22% (95% confidence interval (CI) = −28%, −15%; P = 10^{-9}) 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 (MYBRC; 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 included menarche data from BBCS cases and controls (n = 4,081) in this analysis. 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 ancestry. The mean (± standard deviation) 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).
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 (\( P_{\text{trend}} \)) = 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 (\( P_{\text{het}} \) = 0.44, \( I^2 \) = 1.9% and \( P_{\text{het}} \) = 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 1,114 cases 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 (\( P_{\text{het}} \) = 0.51).

Stratifying cases by oestrogen receptor (\( P_{\text{het}} \) = 0.83) or progesterone receptor (\( P_{\text{het}} \) = 0.19) status, tumour grade (\( P_{\text{het}} \) = 0.63) or nodal involvement at diagnosis (\( P_{\text{het}} \) = 0.51) showed no evidence of effect modification (Table 1). There was some evidence of effect modification by morphology (\( P_{\text{het}} \) = 0.03). For ductal cancers we estimated a very modest reduction of risk for heterozygotes (OR$_{\text{het}}$ = 0.98, 95% CI = 0.93, 1.02; \( P \) = 0.30) and a stronger, significant reduction for homozygotes (OR$_{\text{hom}}$ = 0.74, 95% CI = 0.61, 0.90; \( P \) = 0.003). For lobular cancers there was no such trend (OR$_{\text{het}}$ = 1.07, 95% CI = 0.98, 1.17; \( P \) = 0.14 and OR$_{\text{hom}}$ = 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 RPC) 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 ($P_{\text{het}} = 0.89$; 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% ($OR_{\text{het}} = 0.84, 95\% \text{ CI} = 0.75, 0.94; P = 0.003$) or 19% ($OR_{\text{hom}} = 0.81, 95\% \text{ CI} = 0.51, 1.30; P = 0.39$) ($P_{\text{trend}} = 0.002$) 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 ($OR_{\text{het}} = 1.06, 95\% \text{ CI} = 0.95, 1.19; P = 0.30$ and $OR_{\text{hom}} = 1.07, 95\% \text{ CI} = 0.67, 1.72; P = 0.78$) ($P_{\text{trend}} = 0.29$). 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 2 rs10235235 and risk of breast cancer for women of European ancestry by age at diagnosis

| Age at diagnosis | Cases* | Controls* | OR$_{het}$ | 95% CI | $P_t$ | OR$_{homo}$ | 95% CI | $P_t$ | $P_{het}$ |
|-----------------|--------|-----------|------------|--------|-------|------------|--------|-------|---------|
| ≤ 50 years      | 11,794 | 34,988    | 0.99       | 0.93, 1.05 | 0.69 | 0.83       | 0.53, 0.86 | 0.003 |         |
| > 50 years      | 23,264 | 28,258    | 0.97       | 0.93, 1.02 | 0.24 | 0.84       | 0.70, 1.00 | 0.04  |         |
| NK              | 554    |           |            |         |       |            |        |       |         |
| Total           | 35,612 | 34,988    | 0.98       | 0.94, 1.02 | 0.23 | 0.79       | 0.67, 0.92 | 0.003 | 0.89    |

*Five studies (ABCFS, MARIE, MEC, MTLGEBCS 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 NBBCS) that had no heterozygote or rare homozygote cases in one of the age stratum were also excluded. $H_0$: null hypothesis; NK: not known; OR$_{het}$: odds ratio comparing rs10235235 AG genotype versus AA genotype; OR$_{homo}$: odds ratio comparing rs10235235 GG genotype versus AA genotype; $P_t$: test of $H_0$ no association between rs10235235 and breast cancer risk; $P_{het}$: test of $H_0$ 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 (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 ($D' \approx 1.0, r^2 < 1.0$), 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

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

| rs10235235 genotype | Cases | Age at menarche (years) | $P_{trend}$ | Controls | Age at menarche (years) | $P_{trend}$ |
|---------------------|-------|------------------------|-------------|----------|------------------------|-------------|
| AA                  | 22,954| 12.83                  |             | 23,383   | 12.95                  |             |
| AG                  | 4,312 | 12.83                  |             | 4,627    | 13.02                  |             |
| GG                  | 207   | 12.83                  |             | 248      | 13.05                  |             |
| Total               | 27,473| 12.83                  | 0.97        | 28,258   | 12.96                  | 0.005       |

$H_0$: null hypothesis; $P_{trend}$: test of $H_0$ no linear trend in age at menarche according to rs10235235 genotype.
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.
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 (U01 CA64967), Northern California Cancer Center (U01 CA69417) and University of Melbourne (U01 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. 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 N6 2001-2413 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 BBCC 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 NHS funding to the NIHR Biomedical Research Centre and the National Cancer Research Network. BCAC is funded by CR-UK (CI287/A131018 and CI287/A12014). Meetings of the BCAP have been funded by the European Union COST programme (BM0606), DFE is a Principal Research Fellow of CR-UK. ES (BIGGS) is supported by NHIR 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 Cancer Society and the Nordic Cancer Union. HERPACC was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DFKZ). GENICA was funded by the Federal Ministry of Education and Export Trade and the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. SCCS is supported by a grant from the National Cancer Institute (CA132839). OBCS was supported by the Queensland Health, Labor and Veterans Affairs, the Queensland Cancer Organizations, the Academy of Finland and by the strategic funding of the University of Eastern Finland. KConFab is supported by grants 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 (DAAD17-01-1-0729), 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 (18B-0037, 38R-0102, 5PB-0018, 10PB-0098) 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/SymBoSytl.

The MARE study was supported by the Deutsche Krebshilfe e.V. (70-2892-81), the Humborg Cancer Society, the German Cancer Research Center and the genotype work in part by the Federal Ministry of Education and Research (BMBF) Germany (01HH04022). MBCSG was funded by grants from Italian Association for Cancer Research (AIRC, IG 68731) and by Italian citizens who allocated the 5 x 1000 share of their tax payment in support of the Fondazione I R I C S I n s t i t u t o N a z i o n a l e dei Tumori, according to Italian laws (IN-Istituzional strategic projects 5x1000). MBCE was supported by the NIH grants CA116167 and CA128978, an NIH Specialized Program of Research Excellence (SPRE) 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. GCO was supported by the Dutch Cancer Society (grant number CRN-87221) and the Dutch 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 (UKM/CHIR/MOEHE/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 (BMRC081135/19/550) and the National Medical Research Council, Singapore (NOMRC/CG/SGER/2010). NBCS was supported by grants from the Norwegian Research council, 155218/ V40, 175240/510 to A-LB-D, FUGE-NFR 181600/V11 to VNK and a Swizz 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. OBICS 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 Rausings Initiative Against Breast Cancer. RBCS was funded by the Dutch Cancer Society (DDH4 2004-2124, DDH4 2009-4318). The MARIE study was supported by funding from the European Community’s 7th Framework Programme (SymBioSysII).
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 the 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 U54CA082616-17. SEBCS was supported by the Korea Health 21 R&D Project (A030001), Ministry of Health and Welfare, Republic of Korea. SGBCC is funded by the National Medical Research Council of Thailand. TNBCC 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 Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) – Research Funding Program of the General Secretariat for Research & Technology: ARISTEB. TBWCS 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
1 Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. 2 Division of Breast Cancer Research, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK. 3 Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. 4 Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Way, Belmont, Sutton, Surrey SM2 5NG, UK. 5 The Academic Department of Biochemistry, The Royal Marsden Hospital, Fulham Road, London SW3 6JB, UK. 6 Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, 1-100 Grattan Street, Parkville, Melbourne, Victoria 3010, Australia. 7 Genetic Epidemiology Department, Department of Pathology, The University of Melbourne, 1-100 Grattan Street, Parkville, Melbourne, Victoria 3010, Australia. 8 Division of Molecular Pathology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, P.O. Box 9401, 1066 CX Amsterdam, The Netherlands. 9 Sanquin, Radboud Universiteit Nijmegen, 6525 GA Nijmegen, The Netherlands. 10 Warwick Medical School, University of Warwick, Coventry CV4 7AJ, UK. 11 University Breast Center, Department of Gynecology and Obstetrics, University Hospital Erlangen, Postfach 2306, D-91012 Erlangen, Germany. 12 David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, University of California, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA. 13 Institute of Human Genetics, Friedrich Alexander University Erlangen-Nuremberg, Schloßplatz 4, 91054 Erlangen, Germany. 14 Division 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. 15 Welcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. 16 Oxford Biomedical Research Centre, University of Oxford, The Churchill Hospital, Old Road, Headington OX3 7EL Oxford UK. 17 Surgery, Clinical Science Institute, Galway University Hospital and National University of Ireland, University Road, Galway, Ireland. 18 Department of Obstetrics and Gynecology, University of Heidelberg, Voßstrasse 9, 69115 Heidelberg, Germany. 19 Unit Molecular Epidemiology C880, German Cancer Research Center, DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 20 National Center for Tumor Diseases, University of Heidelberg, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 21 Inserm (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, 101 rue de Tolbiac, Villejuif, 75654 Paris, France. 22 University Paris-Sud, UMR 1018, 101 rue de Tolbiac, Villejuif, 75654 Paris, France. 23 Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Herlev 75, 2730 Herlev, Copenhagen, Denmark. 24 Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev 75, 2730 Herlev, Copenhagen, Denmark. 25 Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev 75, 2730 Herlev, Copenhagen, Denmark. 26 Genetic and Molecular Epidemiology Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Calle de Melchor Fernández Almagro, 3, 28029 Madrid, Spain. 27 Centro de Investigación en Red de Enfermedades Raras (CIBERER), Calle de Melchor Fernández Almagro, 3, 28029 Madrid, Spain. 28 Division of Cancer Etiology, Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA, USA. 29 Department of Epidemiology, School of Medicine, 224 Irvine Hall, University of California Irvine, Irvine, California 92697-7550, USA. 30 Cancer Prevention Center of California, 2201 Walnut Avenue, Suite 300, Fremont, California 95438, USA. 31 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Im Neuenheimer Feld 280, 69121 Heidelberg, Germany. 32 German Cancer Consortium (DKTK), Im Neuenheimer Feld 280, 69121 Heidelberg, Germany. 33 Clinic of Gynaecology and Obstetrics, Division of Tumor Genetics, Klinikum rechts der Isar, Technical University Munich, Ismaninger Strasse 22, D-81675 Munich, Germany. 34 Institute of Human Genetics, University of Heidelberg, Im Neuenheimer Feld 366, 69121 Heidelberg, Germany. 35 Division of Molecular Gynecology-Oncology, Department of Gynaecology and Obstetrics, Center of Molecular Medicine Cologne (CMMC), University Hospital of Cologne, ZMKK-Forschungsgebäude, Robert-Koch-Strasse 21, 50931 Cologne, Germany. 36 Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Robert Bosch Stiftung GmbH, Heidelbergstrasse 31, 70184 Stuttgart, Germany. 37 University of Tübingen, Geschwister-Scholl-Platz, 72074 Tübingen, Germany. 38 Division of Internal Medicine, Evangelische Kliniken Bonn Gmbh, Jochaniter Krankenhaus, 53113 Bonn, Germany. 39 Department of Obstetrics and Gynecology, Helsinki University Central Hospital, University of Helsinki, P.O. Box 140 Haartmanninkatu 2, FIN-00029 Helsinki, Finland. 40 Department of Clinical Genetics, Helsinki University Central Hospital, P.O. Box 140 Haartmanninkatu 2, FIN-00029 Helsinki, Finland. 41 Department of Oncology, Helsinki University Central Hospital, P.O. Box 140 Haartmanninkatu 2, FIN-00029 Helsinki, Finland. 42 Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, 1-1 Nakakodan, Chikusa-ku, Nagoya 464-8681, Japan. 43 Department of Obstetrics and Gynaecology, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. 44 N.N. Alexandrov Research Institute of Oncology and Medical Radiology, 223040, p. Lesnoy, Minsk, Belarus. 45 Department of Molecular Medicine and Surgery, Karolinska Institutet, Solnavägen 1, 171 77 Solna, Stockholm, Sweden. 46 School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Yliopistonranta 1, P.O. Box 1627, FI-70211 Kuopio, Finland. 47 Biocenter Kuopio, Cancer Center of Eastern Finland, University of Eastern Finland, Yliopistonranta 1, P.O. Box 1627, FI-70211 Kuopio, Finland. 48 Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, P.O. Box 100, FI-70209 Kuopio, Finland. 49 Department of Genomics, Queensland Institute of Medical Research, 300 Herston Rd, Herston, Brisbane Queensland 4006, Australia. 50 Department of Preventive Medicine, Keck School of Medicine,
References

1. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001, 358:389–399.
2. Lichtenstein P, Holm NV, Verkasalo PK, Lilja A, Kaprio J, Koskenvuo M, Pukkala E, Skutee A, Hemminki K. Environmental and heritable factors in the causation of cancer—analyses of cohorts from twins of Sweden, Denmark, and Finland. N Engl J Med 2000, 343:78–85.
3. Petro J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. Nat Genet 2000, 26:411–414.
4. Pharoah PD, Aitison J, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002, 31:13–36.
5. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Cl audio J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahmann N, Fletcher Q, Petro J, Gibson L, Los Santos Silva I, Nevanlinna H, Muren A, At etomiaki K, Blomqvist C, Czene K, Iwamoto A, Liu J, Wainsch Q, Meijers-Heijboer H, Adank M, van den Luijt RJ, et al. Large- scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013, 45:353–361.
6. García-Closas M, Couch FF, Lintbrock F, Michalidou K, Schmidt MK, Brook MO, Rør N, Rhië SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, Miron P, Fasching PB, Brauh H, Chang-Cl audie J, Carpenter J, Goody AK, Nevanlinna H, Giles GG, Cox A, Hoppler J, Bolla MK, Wang Q, Dennis J, Dicks E, Howat WJ, Schoof N, Bojesen SE, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013, 45:392–398.
7. Dunning AM, Dowsett M, Healey CS, Tee L, Luben RN, Folkard E, Novik KL, Kelemen L, Ogata S, Pharoah PD, Easton DF, Day NE, Ponder BA. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. J Natl Cancer Inst 2009, 96:636–645.
8. Key T, Appleby P, Barnes J, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 2002, 94:606–616.
9. Walker K, Bratton DJ, Frost C. Premenopausal endogenous oestrogen levels and breast cancer risk: a meta-analysis. Br J Cancer 2011, 105:1451–1457.
10. Fartoukh R, Eklund AH, Spiegelman D, Willett WC, Barbieri RL, Hankinson SE. Premenopausal endogenous steroid hormones and breast cancer risk: results from the Nurses’ Health Study II. Breast Cancer Res 2013, 15:R19.
11. Helzlsouer KJ, Alberg AJ, Bush TL, Longcope C, Gordon GB, Comstock GW. Premenopausal estradiol and breast cancer risk: a new method of controlling for day of the menstrual cycle. Am J Epidemiol 1997, 122:791–799.
12. Rosenberg CR, Pasternack BS, Luber RN, Folkard E, Novik KL, Koenig KL, Toniolo PG. Premenopausal estradiol levels and the risk of breast cancer: a new prospective study of endogenous serum hormone concentrations and breast cancer risk in premenopausal women on the island of Guernsey. Br J Cancer 1997, 75:1075–1079.
13. Kabuto M, Akiha S, Stevens RG, Nerishi K, Land CE. A prospective study of estradiol and breast cancer in Japanese women. Cancer Epidemiol Biomarkers Prev 2000, 9:575–579.
14. Kaaks R, Berinno F, Key T, Rinaldi S, dos Santos Silva, Berrino F, Key T, Bingham S, Boeing H, de Meirleir K, Clavel-Chapelon F, Fournier A, van Gilis CH, Gonzalez CA, Gurea AB, Crissleau J, Khaw KT, Krogh V, Lahi man P, Nagel G, Olsen A, Orlandi Moret NC, Overvad K, Paill D, Panico S, Peeters P, Quiros JR, Raddum A, et al. Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). J Natl Cancer Inst 2005, 97:755–765.
15. Eklund AH, Missmer SA, Tworoger SS, Spiegelman D, Barbieri RL, Dowsett M, Hankinson SE. Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. J Natl Cancer Inst 2006, 98:1406–1415.
16. Stone J, Folkert E, Doody D, Schroen C, Treloar SA, Giles GG, Pike MC, English DR, Soutby MC, Hopper JL, Dowsett M. Familial correlations in postmenopausal serum concentrations of sex steroid hormones and other mitogens: a twins and sisters study. J Clin Endocrinol Metab 2009, 94:4793–4800.
17. Johnson N, Walker K, Gibson L, Orr N, Folkert E, Haynes B, Palles C, Coupland B, Schoemaker M, Jones M, Broderick P, Sawyer E, Kerin M, Tomlinson I, Zvelebil M, Chicolz-Burns S, Tomczyk K, Simpson G, Williamson J, Hillier SG, Ross G, Houlston RS, Sworder A, Ashworth A, Dowsett M, Petro J, Los Santos Silva I, Fletcher C. CYTP3A variation, premenopausal estrogen levels, and breast cancer risk. J Natl Cancer Inst 2012, 104:657–669.
18. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003, 327:557–560.
19. Dvornik V, Wang R H. Genetics of age at menarche: a systematic review. Hum Reprod Update 2012, 18:198–210.
20. He C, Murabito JM. Genome-wide association studies of age at menarche and age at natural menopause. Mol Cell Endocrinol 2014, 382:767–779.
21. Colditz GA, Barr H, Tamimi RM. Breast cancer. In Cancer Epidemiology and Prevention. 3rd edition. Edited by Schottenfeld D, Fraumeni J F. New York: Oxford University Press Inc; 2006:995–1011.
22. Precot J, Thompson DJ, Kolf P, Chalak SI, Audley T, Brown J, Leyland J, Folkert E, Doody D, Hankinson SE, Hunter DJ, Jacobs KB, Dowsett M, Cox DG, Easton DF, De Voo I. Genome-wide association study of circulating estradiol, testosterone, and sex hormone-binding globulin in postmenopausal women. PLoS One 2012, 7e37815.