Breast Cancer
R E S E A R C H

Open Access

Genetic variation in the estrogen metabolic pathway and mammographic density as an intermediate phenotype of breast cancer

Jingmei Li[1,2*], Louise Eriksson[1], Keith Humphreys[1], Kamila Czene[1], Jianjun Liu[2], Rulla M Tamimi[3,4], Sara Lindström[5], David J Hunter[3], Celine M Vachon[6], Fergus J Couch[7], Christopher G Scott[6], Pagona Lagiou[4,8], Per Hall[1]

Abstract

Introduction: Several studies have examined the effect of genetic variants in genes involved in the estrogen metabolic pathway on mammographic density, but the number of loci studied and the sample sizes evaluated have been small and pathways have not been evaluated comprehensively. In this study, we evaluate the association between mammographic density and genetic variants of the estrogen metabolic pathway.

Methods: A total of 239 SNPs in 34 estrogen metabolic genes were studied in 1,731 Swedish women who participated in a breast cancer case-control study, of which 891 were cases and 840 were controls. Film mammograms of the medio-lateral oblique view were digitalized and the software Cumulus was used for computer-assisted semi-automated thresholding of mammographic density. Generalized linear models controlling for possible confounders were used to evaluate the effects of SNPs on mammographic density. Results found to be nominally significant were examined in two independent populations. The admixture maximum likelihood-based global test was performed to evaluate the cumulative effect from multiple SNPs within the whole metabolic pathway and three subpathways for androgen synthesis, androgen-to-estrogen conversion and estrogen removal.

Results: Genetic variants of genes involved in estrogen metabolism exhibited no appreciable effect on mammographic density. None of the nominally significant findings were validated. In addition, global analyses on the overall estrogen metabolic pathway and its subpathways did not yield statistically significant results.

Conclusions: Overall, there is no conclusive evidence that genetic variants in genes involved in the estrogen metabolic pathway are associated with mammographic density in postmenopausal women.

Introduction

Mammographic breast density is one of the strongest risk factors for breast cancer. Several studies have shown that women with extensive dense tissue are at two to six times higher risk of developing the disease than women of similar age with lower mammographic density [1,2]. A strong genetic basis has been suggested for mammographic density [3]. Twin studies have estimated the heritability of this trait to be between 60 and 67% [4]. Evidence for a genetic influence also comes from other studies on family history, familial aggregation and segregation analyses [5,6].

Mammographic density is strongly correlated with hormone exposure profiles of women [7]. Several hormonal risk factors for breast cancer have been found to influence mammographic density in a similar fashion to their respective associations with risk for the disease [8]. For example, a strong inverse relationship has been observed between parity on mammographic density [9]. In addition, hormone replacement therapy (HRT) users and women who have a late first-born child or late menopause have higher breast densities on average [9]. In view of evidence suggesting an association between mammographic density and hormone-related factors, and the fact that estrogen is a strong risk factor for postmenopausal breast cancer, efforts have been made to identify underlying genetic determinants of

* Correspondence: Jingmei.Li@ki.se
[1]Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, 171 77 Stockholm, Sweden

BioMed Central

mammographic density within pathways related to steroid hormone biosynthesis and metabolism [10-13]. Such endeavors assume mammographic density to be an intermediate phenotype for breast cancer. Several genes involved in hormone-related pathways - such as HSD3B1 [5,14], COMT [11,14] and ESR1 [15] - have been suggested to be associated with mammographic breast density. Findings are inconsistent, however, and only few candidate genes have been studied at a time.

We recently reported the results of a study evaluating a total of 239 SNPs in 34 estrogen metabolic genes in 1,596 breast cancer cases and 1,730 population controls from Sweden, of which the outcome variable was breast cancer (Low et al., manuscript submitted). No significant SNP association was evident after correction for multiple testing, but pathway-based global tests revealed significant association evidence for the overall estrogen metabolic pathway (P = 0.034) and, in particular, the androgen-to-estrogen conversion subpathway (P = 0.007). In the present study, we comprehensively examine genetic variation in the estrogen metabolic pathway with mammographic density. The number of SNPs and genes studied provides the most extensive coverage to date with respect to studying mammographic breast density.

Materials and methods
Study subjects
The subjects included in the current study are drawn from a population-based case-control study of postmenopausal breast cancer in women born in Sweden aged 50 to 74 years at the time of enrollment, which was between 1 October 1993 and 31 March 1995. Controls were randomly selected from the Swedish Total Population Register and were frequency matched to the expected age distribution of the cases. Details on data collection and subjects have been described previously [16]. The final study group with both mammographic density and genotype data included 891 breast cancer cases and 840 controls. Although all women were postmenopausal at the time of recruitment to the parent study, a subset of the women (43/1,731) was premenopausal in reference to the date of mammogram.

Approval of the study was given by the ethical review board at the Karolinska Institutet (Stockholm, Sweden) and six other ethical review boards in the respective regions in which the subjects were based, and informed consent was obtained from each participant.

Validation of SNPs with significant associations was performed using mammographic density data from two other studies.

Mammographic density data
The process of collecting mammographic density data in this study has been described previously [17]. Film mammograms of the medio-lateral oblique view were digitized using an Array 2905HD Laser Film Digitizer (Array Corporation, Tokyo, Japan), which covers a range of 0 to 4.7 optical density. For controls, the breast side was randomized. For cases, the side contralateral to the tumor was used. The density resolution was set at 12-bit spatial resolution. The Cumulus software used for the computer-assisted measurement was developed at the University of Toronto [18]. For each image, a trained observer (LE) set the appropriate gray-scale threshold levels defining the edge of the breast and distinguishing dense from nondense tissue. The software calculated the total number of pixels within the entire region of interest and within the region identified as dense. These values were used to calculate the percentage of the breast area that is dense. A random 10% of the images were included as replicates to assess the intra-observer reliability, which was high with a Spearman rank correlation coefficient of 0.95.

Gene and SNP selection
The process of gene and SNP selection has been described in detail by Low et al. (manuscript submitted). A total of 1,007 SNPs were selected from 35 genes and their 30 kb flanking sequences that code the enzymes involved in estradiol or estrone metabolism and are expressed in the breast. These SNPs were genotyped in 92 Swedish control samples to assess linkage disequilibrium patterns, to select tagging SNPs (tagSNPs) and to evaluate their coverage.

Haplotypes were reconstructed using the partition-ligation-expectation-maximization algorithm [19] implemented in the tagSNPs program [20]. A subset of tagSNPs were selected based on the R^2 coefficient, which quantifies how well the tagSNP haplotypes predict the genotype or the number of copies of haplotypes an individual carries. The performance of tagSNPs in capturing unobserved SNPs within the genes was evaluated using a SNP-dropping analysis. In brief, each of the genotyped SNPs was dropped in turn and then tagSNPs were selected from the remaining SNPs so that their haplotypes predicted the remaining SNPs with an R^2 value of 0.85. In total, 312 tagSNPs from the 35 genes were selected for genotyping.

Figure 1 delineates the processes and genes involved in the androgen synthesis, androgen-to-estrogen conversion and estrogen removal subpathways. The lists of SNPs corresponding to each subpathway are summarized in Tables S1 to S3 in Additional file 1.

DNA extraction and genotyping
DNA was extracted from 4 ml whole blood using the QIAamp DNA Blood Maxi Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions
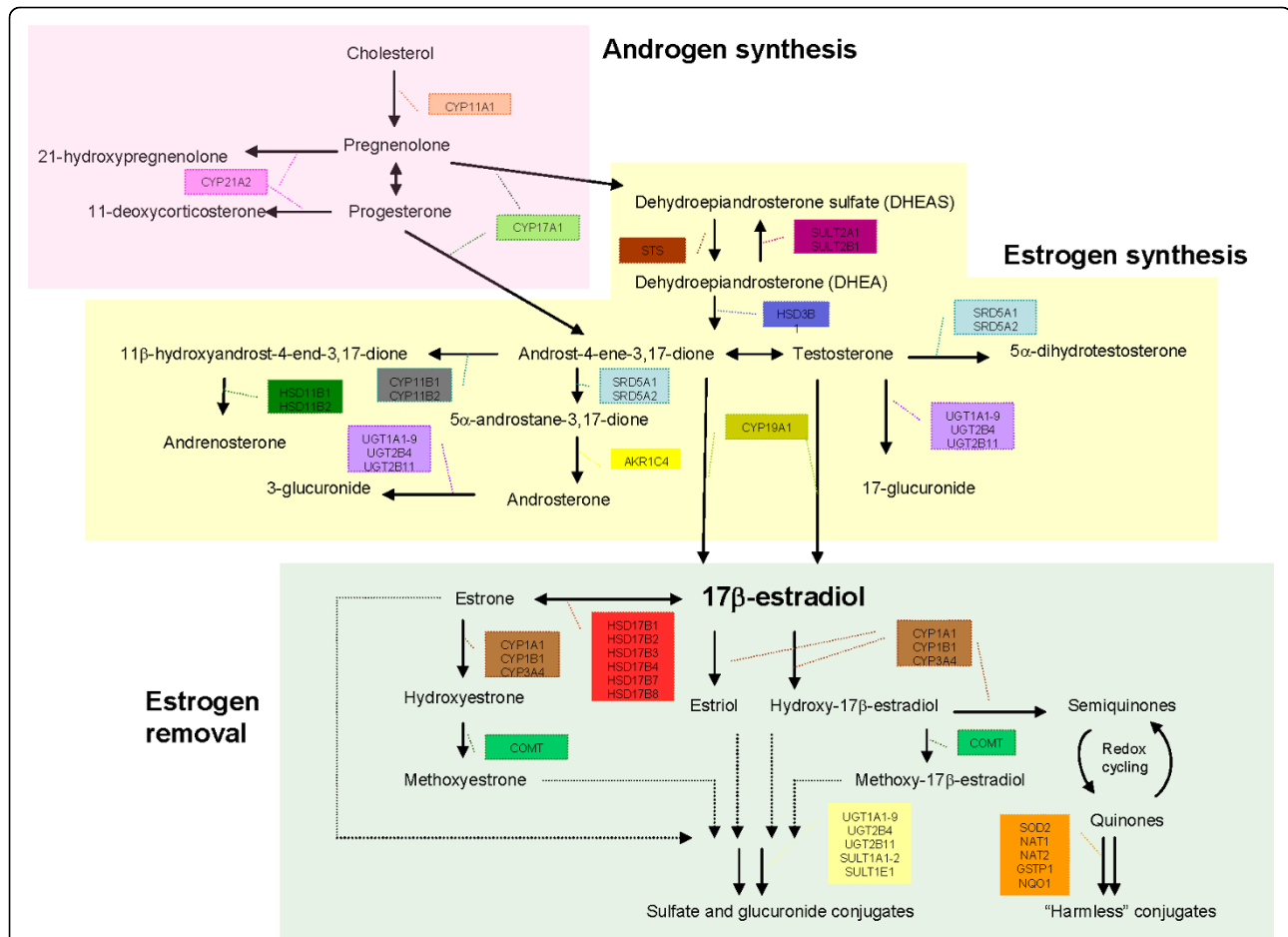
Figure 1 Subdivision of the estrogen metabolic pathway. The 34 metabolic genes analyzed in the present study are involved in different steps of the estrogen metabolism. The genes are divided into the three groups involved in androgen synthesis, estrogen synthesis and estrogen removal for subpathway-based association analysis.

and nonmalignant cells in paraffin-embedded tissue using a standard phenol/chloroform/isoamyl alcohol protocol. Genotyping was performed using the primer extension-based assay from Sequenom (San Diego, CA, USA) according to the manufacturers' instructions. DNA samples were randomly assigned to the plates carrying positive and negative controls, and all genotyping results were generated and checked by laboratory staff unaware of the case-control status. SNPs with a call rate <85%, minor allele frequency <1% or out of Hardy-Weinberg equilibrium ($P < 0.05/312$) were excluded from further analysis. The genotype concordance was >99%, suggesting high genotyping accuracy. Overall, 239 tagSNPs from the 34 genes were successfully genotyped and used in statistical analysis.

Statistical analysis

Linear regression models were fitted, treating percentage density as an outcome. Models were adjusted for age, body mass index, menopausal status and HRT. Age was coded as 0, 1 and 2 for women <50 years, between 50 and 60 years, and >60 years of age, respectively. The body mass index was treated as a continuous variable. Menopausal status was determined from the time difference between the date of menopause and the date on which the mammogram was taken. HRT was considered a categorical variable made up of three groups: never users, past users and current users. The mammographic density measurements were transformed by the power of 0.3, yielding an approximately normal distribution. The genotypes were coded 0, 1 and 2 and treated as continuous variables.

A likelihood ratio test was performed for each SNP. Normal quantile-quantile plots were used to examine the distributions of the -$\log_{10}$-transformed P values. To assess whether the SNPs associated with breast cancer risk are the same SNPs as those associated with mammographic density, we used the Spearman's rank correlation test, evaluating the relationship between odds ratios corresponding to SNP effects on breast cancer

risk and the regression coefficients of SNP effects on percentage density. The admixture maximum likelihood-based global test [21] was performed to evaluate the cumulative effect on mammographic density from multiple SNPs within the whole metabolic pathway and three subpathways for androgen synthesis, androgen-to-estrogen conversion and estrogen removal. Affection status for the admixture maximum likelihood analysis was defined by taking the lowest quantile of all percentage density measurements as controls and the highest quantile as cases. *P* values of the admixture maximum likelihood test were obtained via 5,000 permutations. Software R (v2.8.0) [22] and admixture maximum likelihood [21] were used for data management, quality control and statistical analyses.

Validation of significantly associated SNPs

SNP associations with mammographic density were validated in 1,590 women genotyped with the Illumina HumanHap500 as part of the Cancer Genetic Markers of Susceptibility Project (CGEMS) [23]. The CGEMS project is a National Cancer Institute initiative to conduct genome-wide association studies to identify genes involved in breast cancer and prostate cancer. The initial CGEMS breast cancer scan was designed and funded to study the main effect of SNP variants on breast cancer risk in postmenopausal women, and has been completed [24]. Briefly, the first stage of the project involved a whole genome scan of 1,145 invasive postmenopausal breast cancer cases and 1,142 matched controls from the Nurses' Health Study nested case-control study [24]. The Nurses' Health Study was initiated in 1976, when 121,700 US registered nurses aged 30 to 55 returned an initial questionnaire [25]. During 1989 and 1990, blood samples were collected from 32,826 women [26]. For 1,590 of these women - of which 806 were breast cancer cases and 784 were healthy controls - we also had mammographic density measurements.

We collected mammograms as close as possible to the date of blood collection (1989 to 1990). To assess mammographic density, the craniocaudal (CC) views of both breasts were digitized at 261 µm/pixel with a Lumysis 85 laser film scanner, which covers a range of 0 to 4.0 optical density. The software for computer-assisted thresholding was developed at the University of Toronto [18]. We used the average percentage density of both breasts for this analysis. This collection has been described in detail in a previous publication [27]. SNPs not available on the Illumina HumanHap550 panel were imputed using MACH [28] based on HapMap Phase II (release 21a). For the analysis of imputed data, the ProbABEL package from the ABEL set of programs was used [29]. Percentage density was transformed by the power of 0.3 to be consistent with the parent study.

This study was approved by the Committee on the Use of Human Subjects in Research at Brigham and Women's Hospital.

The second validation population consisted of a set of controls from an ongoing breast cancer case-control study at the Mayo Clinic. Briefly, the Mayo Clinic Breast Cancer Study is an Institutional Review Board-approved, clinic-based, case-control study initiated in February 2001 at Mayo Clinic, Rochester, MN, USA. The study design has been presented previously [30,31]. Clinic attendance formed the sampling frame for Mayo Clinic cases and controls. Consecutive cases were women aged 18 years or over with histologically confirmed primary invasive breast carcinoma and recruited within 6 months of the date of diagnosis. Cases lived in the six-state region that defines Mayo Clinic's primary service population (Minnesota, Iowa, Wisconsin, Illinois, North Dakota, and South Dakota). Controls without prior history of cancer (other than nonmelanoma skin cancer) were frequency matched on age (5-year age category), race and six-state region of residence to cases. Controls were recruited from the outpatient practice of the Divisions of General Internal Medicine and Primary Care Internal Medicine at Mayo Clinic, where they were seen for routine medical examinations.

The present analysis genotyped Caucasian controls (99% of study participants) enrolled through September 2007, who had mammograms available, representing 995 total controls (76% of total possible controls), of which 783 were postmenopausal. Screening mammograms were ascertained close to the enrollment date and the left CC view was digitized on an Array 2905HD Laser Film Digitizer, which covers a range of 0 to 4.7 optical density. Percentage mammographic density was estimated by an expert reader [32] on the left CC view, using the same Cumulus software described above [33]. Genotyping was carried out using TaqMan (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions, using 10 to 20 ng DNA. Primers and probes were Assay-by Design (Applied Biosystems). Following PCR amplification, end reactions are read on the ABI Prism 7900 ht using Sequence Detection Software (Applied Biosystems). SNP associations were examined only in the 783 postmenopausal controls, to be comparable with the two other populations. The percentage density was transformed by the power of 0.3 to be consistent with the parent study.

Results

Our dataset consisted of 1,731 postmenopausal women, of which 981 were breast cancer cases and 840 were controls (Table 1). Cases and controls differed significantly in age at first birth ($P = 0.0126$), parity ($P < 0.0001$), family history of breast cancer ($P = 0.0002$) and

Table 1 Selected characteristics of subjects

	Breast cancer cases (n = 891)		Breast cancer controls (n = 840)		
	Mean	SD	Mean	SD	P value
Age (years)	63.0	6.3	63.0	6.3	0.9045
Height (cm)	164.1	5.7	163.6	5.5	0.0766
Weight (kg)	68.9	110	68.8	11.6	0.8153
Body mass index	25.6	3.9	25.6	4.1	0.8420
Age at menarche (years)	13.6	1.4	13.6	1.5	0.6090
Age at first birth (years)	25.4	50	24.8	4.7	0.0126
Parity	1.9	1.2	2.2	1.3	0.0000
Age at menopause (years)	50.3	3.6	50.1	3.9	0.1223
HRT (% ever use)	0.53		0.50		0.2523
Family history (%)	0.15		0.09		0.0002
Percent density	16.7	14.3	14.6	14.0	0.0017

Means and standard deviations (SD) are given for continuous measures, proportions for other variables. *P* values based on the Welch *t*test for independent samples. HRT, hormone replacement therapy.

percentage density (*P* = 0.0017). Cases were found to have higher percentage density (mean ± standard deviation: 16.7 ± 14.3) than controls (14.6 ± 14.0). No significant difference was found for age, height, weight, body mass index, age at menarche, age at menopause or HRT usage.

Table S4 in Additional file 2 shows a list of 34 genes involved in the estrogen metabolic pathway and the corresponding number of SNPs examined for each gene. References are given for genes that have been examined in other studies for an association with mammographic density. Of the 239 SNPs analyzed, 11 SNPs were found to be significant at the 5% level (Table 2) - of which the smallest *P* value was 0.0019. Among six tagSNPs selected for the gene CYP11A1, five were found to be significant in the same direction. The associations in the single SNP analysis were moderate and would not survive correction for multiple SNP testing. In addition, the single-SNP *P* values showed no clear deviation from the null distribution, representing no association between SNPs and percentage density (Figure 2; see also Tables
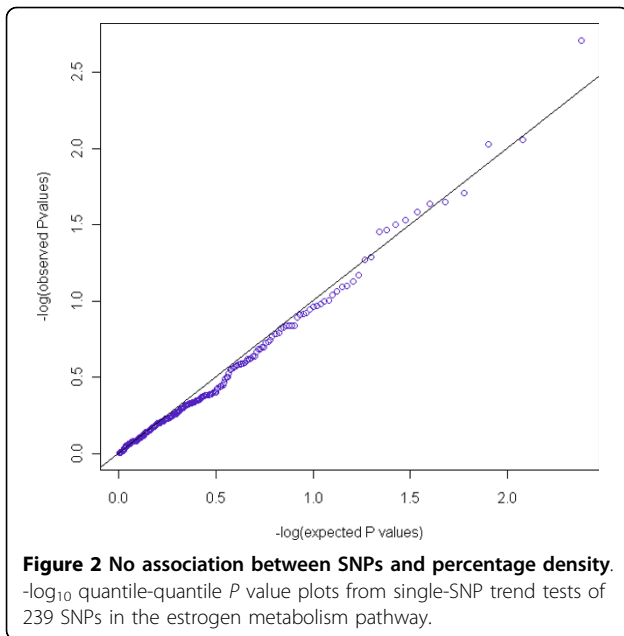
S1 to S3 in Additional file 1). None of the SNPs found to be nominally significant in our dataset were found to be significant in the CGEMS validation set (see Table S5 in Additional file 3). A second, independent validation carried out on the most significantly associated SNP (rs11638442) located within the CYP11A1 gene in 783 postmenopausal women with mammograms in the Mayo Clinic Breast Cancer Study yielded a *P* value of 0.88 (regression coefficient = -0.000507, 95% confidence interval = -0.07251 to 0.06237).

Since the estrogen metabolic SNPs examined have previously been associated with breast cancer risk, we estimated the correlation between regression coefficients of SNP effects on mammographic density and the odds ratios of SNP effects on breast cancer risk, in order to assess whether the SNPs act through mammographic density as an intermediate phenotype for breast cancer. No significant relationship was found between SNP effects on breast cancer risk and percentage density (Spearman's correlation rho = 0.0411, *P* = 0.5268). Pathway-based multi-SNP association analyses

Table 2 Significant SNPs in the estrogen metabolic pathway, corresponding regression coefficients and P values

SNP	Gene	Minor allele	MAF	n	Coefficient	SE	P value
rs11638442	CYP11A1	C	0.35	1,677	0.0557	0.0212	0.0088
rs16968478	CYP11A1	G	0.17	1,703	0.0575	0.0263	0.0293
rs2279357	CYP11A1	A	0.20	1,699	0.0511	0.0229	0.0260
rs2959003	CYP11A1	A	0.28	1,669	0.0582	0.0224	0.0094
rs2959008	CYP11A1	A	0.30	1,703	0.0475	0.0221	0.0315
rs2066485	HSD17B3	G	0.14	1,703	0.0668	0.0293	0.0230
rs7039978	HSD17B3	A	0.50	1,694	−0.0632	0.0203	0.0019
rs1469908	NQO1	C	0.37	1,695	−0.0472	0.0206	0.0223
rs17268974	STS	A	0.22	1,605	0.0503	0.0238	0.0349
rs2270112	STS	C	0.34	1,686	−0.0485	0.0208	0.0197
rs707762	STS	A	0.40	1,687	0.0435	0.0205	0.0340

P values from a one-degree-of-freedom likelihood ratio test. MAF, minor allele frequency; SE, standard error.

Figure 2 No association between SNPs and percentage density.
$-\log_{10}$ quantile-quantile *P* value plots from single-SNP trend tests of 239 SNPs in the estrogen metabolism pathway.

Table 3 Global genetic association tests between SNPs in the estrogen metabolic pathways and mammographic breast density

Pathway	Number of SNPs	P heterogeneity	P trend[a]
Whole pathway	239	0.840	0.507
Androgen synthesis	11	0.761	0.763
Androgen to estrogen conversion	120	0.587	0.715
Estrogen removal	134	0.834	0.872

[a]*P* values based on 5,000 permutations.

revealed no significant association between percentage density and genetic variations in the overall estrogen metabolic pathway, or any of the related subpathways (Table 3).

Discussion

Our study suggests there is no appreciable effect between genetic variants involved in estrogen metabolism and mammographic density. Neither the overall estrogen metabolic pathway nor the androgen synthesis, androgen-to-estrogen conversion and estrogen removal subpathways were found to be significantly associated with mammographic density. Single SNP markers with significant associations with mammographic density were not validated in two independent datasets.

In view of estrogen exposure being a major risk factor of postmenopausal breast cancer, and mammographic density being associated with several hormone-related factors such as body mass index (increased local estrogen conversion due to increased fatty tissue), HRT, and menopausal status, the estrogen metabolic pathway has been a candidate pathway for the search of genetic variants related to mammographic density. Most of the variants in the candidate breast cancer genes evaluated in previous studies, however, have been concluded to be only weak predictors of mammographic density [10]. Association findings have been both supported and contradicted [3]. As Boyd and colleagues have discussed [34], it is likely that hormone-related factors are responsible for only a small proportion of the wide variation in mammographic density. In addition, genetic variants involved in the estrogen metabolic pathway are generally investigated based on the premise that mammographic density is an intermediate and heritable risk factor of breast cancer [4]. There is, however, accumulating evidence that mammographic density may predispose to breast cancer risk through components largely independent of estrogen metabolism [35-37].

In our study, no correlation was found between the estimates of SNP effects on breast cancer risk and mammographic density, suggesting that the same SNPs associated with breast cancer risk are not directly correlated with mammographic density. Tamimi and colleagues reported that mammographic density and circulating sex steroid levels were independently associated with breast cancer risk in postmenopausal women [35]. In addition, Kerlikowske and colleagues found no correlation between mammographic density and bone mineral density [36], both of which have been suggested to be cumulative markers of elevated estrogen exposure. Dite and colleagues performed a similar study investigating the overlap between genetic determinants of mammographic density and bone mineral density, and reported a null finding [37]. Another finding in Kerlikowske and colleagues' study was that although mammographic density remained strongly associated with elevated breast cancer risk after adjustment for hormone-related factors, the effects of bone mineral density did not [36], suggesting that estrogen metabolism plays only a small role in the effects of mammographic density on breast cancer risk.

Many studies examining the effects of exogenous estrogen exposure are in agreement with the view that estrogen has limited effects on mammographic density. Very often, the combined estrogen plus progestin regimen was found to affect mammographic density more than the estrogen-only regimen [38-41], suggesting that progestins and not estrogens are responsible for increased mammographic density. Interestingly, mammographic density is also known to have no prognostic bearing on the estrogen receptor status of breast cancer tumors [42-44], thus corroborating an estrogen/estrogen receptor independent link. Another study conducted by Vachon and colleagues found no

influence of aromatase inhibitors (drugs that stop the production of estrogen in postmenopausal women) on mammographic density [45], further supporting this line of rationale.

Strengths of the present study include the large sample size and extensive coverage of SNPs in the estrogen metabolic pathway. In a review by Kelemen and colleagues, the authors summarized that previous genetic association studies exploring the relationship between the estrogen metabolic pathway and mammographic density had sample sizes ranging from between 232 and 1,260 women [3]. The number of loci involved in the estrogen metabolic pathway investigated in these studies was also limited to eight or less [3], while we examined 239 tagSNPs from 34 genes involved in the estrogen metabolic pathway. A second strength of the present study is the use of two independent populations for the validation of the associations found.

A limitation of the present work is that it includes different mammogram views across the different studies. The main study on Swedish women utilized the medio-lateral oblique view, while mammograms of the CGEMS and of the Mayo Clinic were taken using the CC view. Several studies, however, have shown correlation of densities from the medio-lateral oblique and CC views [46,47], and have shown that the different views yield similar associations with breast cancer [32]. In addition, the main focus of this study was on genetic determinants of mammographic density in postmenopausal women. Although no strong association was observed between SNPs in the estrogen metabolic pathway examined and mammographic density in postmenopausal women, whether the same lack of association between common genetic variation in the estrogen metabolism pathway and mammographic density is present in premenopausal women remains to be clarified.

Conclusions

As mammographic density is generally considered an intermediate phenotype for breast cancer, the identification of genes that influence mammographic density would play an important role in risk prediction of breast cancer prior to the start of mammography screenings and shed light on the mechanisms behind breast cancer carcinogenesis. Overall, there is no conclusive evidence that genetic variants in genes involved in the estrogen metabolic pathway are associated with mammographic density in postmenopausal women. This knowledge will be helpful for directing the focus of future studies to alternative pathways that may be responsible for a larger bulk of the genetic component of mammographic density.

Additional file 1: Tables S1 to S3. Table S1 presents a list of SNPs in the androgen synthesis subpathway and their corresponding regression coefficients and likelihood ratio test *P* values. Table S2 presents a list of SNPs in the androgen to estrogen conversion subpathway and their corresponding regression coefficients and likelihood ratio test *P* values. Table S3 presents a list of SNPs in the estrogen removal subpathway and their corresponding regression coefficients and likelihood ratio test *P* values.
Click here for file
[http://www.biomedcentral.com/content/supplementary/bcr2488-S1.DOC]

Additional file 2: Table S4. Table S4 presents genes containing polymorphisms within the estrogen metabolic pathway evaluated in relation to mammographic density.
Click here for file
[http://www.biomedcentral.com/content/supplementary/bcr2488-S2.DOC]

Additional file 3: Table S5. Table S5 presents validation results of significantly associated SNPs in the Nurses' Health Study (NHS) and the Mayo Clinic Breast Cancer Study (MBCS).
Click here for file
[http://www.biomedcentral.com/content/supplementary/bcr2488-S3.DOC]

Abbreviations

CC: craniocaudal; CGEMS: Cancer Genetic Markers of Susceptibility Project; HRT: hormone replacement therapy; SNP: single nucleotide polymorphism; tagSNP: tagging single nucleotide polymorphism.

Author details

[1]Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, 171 77 Stockholm, Sweden. [2]Human Genetics, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore. [3]Channing Laboratory, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115, USA. [4]Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. [5]Program in Genetic and Molecular Epidemiology, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. [6]Department of Health Sciences Research, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, USA. [7]Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, USA. [8]Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, 75 Mikras Asias Str, Goudi, Athens 115 27, Greece.

Authors' contributions

JLi participated in the study design, carried out the analyses and drafted the manuscript. LE digitized and obtained readings for the mammograms. RMT, SL, DJH, CMV, FJC and CGS contributed to the validation of this study. PL coordinated the Innovator project which contributed data on birthweight and mammographic density. JLi, KH, KC, JLiu and PH conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

References

1. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ: **Mammographic densities and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 1998, **7**:1133-1144.
2. Vachon CM, van Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR, Pankratz VS: **Mammographic density, breast cancer risk and risk prediction.** *Breast Cancer Res* 2007, **9**:217.
3. Kelemen LE, Sellers TA, Vachon CM: **Can genes for mammographic density inform cancer aetiology?.** *Nat Rev Cancer* 2008, **8**:812-823.
4. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, Giles GG, Tritchler D, Chiarelli A, Yaffe MJ, Hopper JL: **Heritability of mammographic density, a risk factor for breast cancer.** *N Engl J Med* 2002, **347**:886-894.
5. Stone J, Gurrin LC, Byrnes GB, Schroen CJ, Treloar SA, Padilla EJ, Dite GS, Southey MC, Hayes VM, Hopper JL: **Mammographic density and candidate gene variants: a twins and sisters study.** *Cancer Epidemiol Biomarkers Prev* 2007, **16**:1479-1484.
6. Boyd NF, Martin LJ, Rommens JM, Paterson AD, Minkin S, Yaffe MJ, Stone J, Hopper JL: **Mammographic density: a heritable risk factor for breast cancer.** *Methods Mol Biol* 2009, **472**:343-360.
7. Boyd NF, Martin LJ, Yaffe MJ, Minkin S: **Mammographic density: a hormonally responsive risk factor for breast cancer.** *J Br Menopause Soc* 2006, **12**:186-193.
8. Duffy SW, Jakes RW, Ng FC, Gao F: **Interaction of dense breast patterns with other breast cancer risk factors in a case-control study.** *Br J Cancer* 2004, **91**:233-236.
9. Martin LJ, Boyd NF: **Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence.** *Breast Cancer Res* 2008, **10**:201.
10. Haiman CA, Hankinson SE, De Vivo I, Guillemette C, Ishibe N, Hunter DJ, Byrne C: **Polymorphisms in steroid hormone pathway genes and mammographic density.** *Breast Cancer Res Treat* 2003, **77**:27-36.
11. Maskarinec G, Lurie G, Williams AE, Le Marchand L: **An investigation of mammographic density and gene variants in healthy women.** *Int J Cancer* 2004, **112**:683-688.
12. Lord SJ, Mack WJ, Berg Van Den D, Pike MC, Ingles SA, Haiman CA, Wang W, Parisky YR, Hodis HN, Ursin G: **Polymorphisms in genes involved in estrogen and progesterone metabolism and mammographic density changes in women randomized to postmenopausal hormone therapy: results from a pilot study.** *Breast Cancer Res* 2005, **7**:R336-R344.
13. Chambo D, Kemp C, Costa AM, Souza NC, Guerreiro da Silva ID: **Polymorphism in CYP17, GSTM1 and the progesterone receptor genes and its relationship with mammographic density.** *Braz J Med Biol Res* 2009, **42**:323-329.
14. Haiman CA, Bernstein L, Berg D, Ingles SA, Salane M, Ursin G: **Genetic determinants of mammographic density.** *Breast Cancer Res* 2002, **4**:R5.
15. van Duijnhoven FJ, Bezemer ID, Peeters PH, Roest M, Uiterlinden AG, Grobbee DE, van Gils CH: **Polymorphisms in the estrogen receptor alpha gene and mammographic density.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:2655-2660.
16. Wedren S, Lovmar L, Humphreys K, Magnusson C, Melhus H, Syvanen AC, Kindmark A, Landegren U, Fermer ML, Stiger F, Persson I, Baron J, Weiderpass E: **Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study.** *Breast Cancer Res* 2004, **6**:R437-R449.
17. Tamimi RM, Eriksson L, Lagiou P, Czene K, Ekbom A, Hsieh CC, Adami HO, Trichopoulos D, Hall P: **Birth weight and mammographic density among postmenopausal women in Sweden.** *Int J Cancer* 126:985-991.
18. Byng JW, Boyd NF, Little L, Lockwood G, Fishell E, Jong RA, Yaffe MJ: **Symmetry of projection in the quantitative analysis of mammographic images.** *Eur J Cancer Prev* 1996, **5**:319-327.
19. Qin ZS, Niu T, Liu JS: **Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms.** *Am J Hum Genet* 2002, **71**:1242-1247.
20. Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC: **Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study.** *Hum Hered* 2003, **55**:27-36.
21. Tyrer J, Pharoah PD, Easton DF: **The admixture maximum likelihood test: a novel experiment-wise test of association between disease and multiple SNPs.** *Genet Epidemiol* 2006, **30**:636-643.
22. R Development Core Team: *R: A Language and Environment for Statistical Computing* Austria: R Foundation for Statistical Computing 2008.
23. Everett BM, Kurth T, Buring JE, Ridker PM: **The relative strength of C-reactive protein and lipid levels as determinants of ischemic stroke compared with coronary heart disease in women.** *J Am Coll Cardiol* 2006, **48**:2235-2242.
24. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Thomas G, Chanock SJ: **A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.** *Nat Genet* 2007, **39**:870-874.
25. Colditz GA, Hankinson SE: **The Nurses' Health Study: lifestyle and health among women.** *Nat Rev Cancer* 2005, **5**:388-396.
26. Hankinson SE, Willett WC, Manson JE, Colditz GA, Hunter DJ, Spiegelman D, Barbieri RL, Speizer FE: **Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women.** *J Natl Cancer Inst* 1998, **90**:1292-1299.
27. Tamimi RM, Cox DG, Kraft P, Pollak MN, Haiman CA, Cheng I, Freedman ML, Hankinson SE, Hunter DJ, Colditz GA: **Common genetic variation in IGF1, IGFBP-1, and IGFBP-3 in relation to mammographic density: a cross-sectional study.** *Breast Cancer Res* 2007, **9**:R18.
28. Li Y, Willer C, Sanna S, Abecasis G: **Genotype imputation.** *Annu Rev Genomics Hum Genet* 2009, **10**:387-406.
29. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM: **GenABEL: an R library for genome-wide association analysis.** *Bioinformatics* 2007, **23**:1294-1296.
30. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, et al: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, **447**:1087-1093.
31. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, Lissowska J, Brinton L, Peplonska B, Southey MC, Hopper JL, McCredie MR, Giles GG, Fletcher O, Johnson N, dos Santos Silva I, Gibson L, Bojesen SE, Nordestgaard BG, Axelsson CK, Torres D, Hamann U, Justenhoven C, Brauch H, Chang-Claude J, Kropp S, et al: **A common coding variant in CASP8 is associated with breast cancer risk.** *Nat Genet* 2007, **39**:352-358.
32. Vachon CM, Brandt KR, Ghosh K, Scott CG, Maloney SD, Carston MJ, Pankratz VS, Sellers TA: **Mammographic breast density as a general marker of breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2007, **16**:43-49.
33. Boyd NF, Stone J, Martin LJ, Jong R, Fishell E, Yaffe M, Hammond G, Minkin S: **The association of breast mitogens with mammographic densities.** *Br J Cancer* 2002, **87**:876-882.
34. Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, Paterson AD: **Mammographic breast density as an intermediate phenotype for breast cancer.** *Lancet Oncol* 2005, **6**:798-808.
35. Tamimi RM, Byrne C, Colditz GA, Hankinson SE: **Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women.** *J Natl Cancer Inst* 2007, **99**:1178-1187.
36. Kerlikowske K, Shepherd J, Creasman J, Tice JA, Ziv E, Cummings SR: **Are breast density and bone mineral density independent risk factors for breast cancer?** *J Natl Cancer Inst* 2005, **97**:368-374.
37. Dite GS, Wark JD, Giles GG, English DR, McCredie MR, Hopper JL: **Is there overlap between the genetic determinants of mammographic density and bone mineral density?** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:2266-2268.
38. Kaewrudee S, Anuwutnavin S, Kanpittaya J, Soontrapa S, Sakondhavat C: **Effect of estrogen-progestin and estrogen on mammographic density.** *J Reprod Med* 2007, **52**:513-520.

39. McTiernan A, Martin CF, Peck JD, Aragaki AK, Chlebowski RT, Pisano ED, Wang CY, Brunner RL, Johnson KC, Manson JE, Lewis CE, Kotchen JM, Hulka BS: Estrogen-plus-progestin use and mammographic density in postmenopausal women: Women's Health Initiative randomized trial. *J Natl Cancer Inst* 2005, **97**:1366-1376.
40. Topal NB, Ayhan S, Topal U, Bilgin T: Effects of hormone replacement therapy regimens on mammographic breast density: the role of progestins. *J Obstet Gynaecol Res* 2006, **32**:305-308.
41. McTiernan A, Chlebowski RT, Martin C, Peck JD, Aragaki A, Pisano ED, Wang CY, Johnson KC, Manson JE, Wallace RB, Vitolins MZ, Heiss G: Conjugated equine estrogen influence on mammographic density in postmenopausal women in a substudy of the women's health initiative randomized trial. *J Clin Oncol* 2009, **27**:6135-6143.
42. Ghosh K, Brandt KR, Sellers TA, Reynolds C, Scott CG, Maloney SD, Carston MJ, Pankratz VS, Vachon CM: Association of mammographic density with the pathology of subsequent breast cancer among postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2008, **17**:872-879.
43. Chen JH, Hsu FT, Shih HN, Hsu CC, Chang D, Nie K, Nalcioglu O, Su MY: Does breast density show difference in patients with estrogen receptor-positive and estrogen receptor-negative breast cancer measured on MRI? *Ann Oncol* 2009, **20**:1447-1449.
44. Ziv E, Tice J, Smith-Bindman R, Shepherd J, Cummings S, Kerlikowske K: Mammographic density and estrogen receptor status of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2004, **13**:2090-2095.
45. Vachon CM, Ingle JN, Suman VJ, Scott CG, Gottardt H, Olson JE, Goss PE: Pilot study of the impact of letrozole vs. placebo on breast density in women completing 5 years of tamoxifen. *Breast* 2007, **16**:204-210.
46. Ursin G, Hovanessian-Larsen L, Parisky YR, Pike MC, Wu AH: Greatly increased occurrence of breast cancers in areas of mammographically dense tissue. *Breast Cancer Res* 2005, **7**:R605-R608.
47. Kim SJ, Moon WK, Cho N, Cha JH, Kim SM, Im JG: Computer-aided detection in digital mammography: comparison of craniocaudal, mediolateral oblique, and mediolateral views. *Radiology* 2006, **241**:695-701.