


RESEARCH ARTICLE

Open Access



Common genetic variation and novel loci associated with volumetric mammographic density

Judith S. Brand^{1*} , Keith Humphreys¹, Jingmei Li^{1,2}, Robert Karlsson¹, Per Hall¹ and Kamila Czene¹

Abstract

Background: Mammographic density (MD) is a strong and heritable intermediate phenotype of breast cancer, but much of its genetic variation remains unexplained.

Methods: We conducted a genetic association study of volumetric MD in a Swedish mammography screening cohort ($n = 9498$) to identify novel MD loci. Associations with volumetric MD phenotypes (percent dense volume, absolute dense volume, and absolute nondense volume) were estimated using linear regression adjusting for age, body mass index, menopausal status, and six principal components. We also estimated the proportion of MD variance explained by additive contributions from single-nucleotide polymorphisms (SNP-based heritability [h^2_{SNP}]) in 4948 participants of the cohort.

Results: In total, three novel MD loci were identified (at $P < 5 \times 10^{-8}$): one for percent dense volume (*HABP2*) and two for the absolute dense volume (*INHBB*, *LINC01483*). *INHBB* is an established locus for ER-negative breast cancer, and *HABP2* and *LINC01483* represent putative new breast cancer susceptibility loci, because both loci were associated with breast cancer in available meta-analysis data including 122,977 breast cancer cases and 105,974 control subjects ($P < 0.05$). h^2_{SNP} (SE) estimates for percent dense, absolute dense, and nondense volume were 0.29 (0.07), 0.31 (0.07), and 0.25 (0.07), respectively. Corresponding ratios of h^2_{SNP} to previously observed narrow-sense h^2 estimates in the same cohort were 0.46, 0.72, and 0.41, respectively.

Conclusions: These findings provide new insights into the genetic basis of MD and biological mechanisms linking MD to breast cancer risk. Apart from identifying three novel loci, we demonstrate that at least 25% of the MD variance is explained by common genetic variation with h^2_{SNP}/h^2 ratios varying between dense and nondense MD components.

Keywords: Mammographic density, Genetic association study, Heritability

Background

Mammographic density (MD) refers to the amount of fibroglandular dense and fatty nondense tissues in the breast, which appear white and black, respectively, on an X-ray mammogram. Women with high MD (i.e., a high amount of fibroglandular dense tissue) are at four to sixfold increased risk of developing breast cancer compared with women having nondense or fatty breasts [1]. Despite being an important determinant of breast cancer risk, the biological mechanisms determining tumor development in women with highly dense breasts are not

well understood. Understanding the genetic basis of MD and how its associated loci are related to breast cancer risk may shed light on the processes that are responsible for the transformation of mammographic dense tissue into tumor tissue. Moreover, as a strong intermediate phenotype for breast cancer [2], genetic association studies of MD have the potential to identify new breast cancer susceptibility loci.

Family-based studies have estimated that approximately 60–70% of the variance in MD is explained by additive genetic effects, which is considerably higher than the narrow-sense heritability estimate reported for breast cancer ($h^2 = 30$ –40%). To date, genome-wide association studies (GWAS) [3–7] have identified MD-associated single-nucleotide polymorphisms (SNPs) at

* Correspondence: judith.brand@ki.se

¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels Väg 12A, 171 77 Stockholm, Sweden

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



13 loci (1q12.21, *AREG*, *PRDM6*, *TAB2*, *CCDC170/ESR1*, *ESR1*, 8p11.23, *ZNF365*, *LSP1*, *IGF1*, 12q24, *TMEM184B*, and *SGSM3/MKL1*) across the genome, most of which have also been associated with breast cancer. However, additive effects of these genome-wide significant SNPs explain only a small fraction of the total MD variance (<5%). This discrepancy in explained variance or “missing heritability” has been attributed to various factors, including the presence of large numbers of common variants with small effects, rare variants with large effects not tagged by common SNPs on genotyping arrays and possible inflation of narrow-sense h^2 estimates if family resemblance is influenced by nonadditive effects, shared environmental effects, and/or gene-environment interactions. Although the exact contribution of each of these possible explanations is difficult to determine, the contribution of common genetic variation to MD variance (SNP-based heritability or $[h^2_{\text{SNP}}]$) can be assessed rather easily in genome-wide association data and can provide insights into the heritability fraction that remains to be identified in future larger-scale GWAS.

In the present study, we conducted a large-scale genetic association analysis of volumetric MD to identify novel MD loci. We further sought to unravel the genetic basis of MD through estimating the proportion of MD variance attributable to common SNPs. Previous GWAS of MD have relied mainly on area-based or 2D MD measures derived from film mammograms using labor-intensive semiautomated methods [8]. For this analysis, we used volumetric MD measures derived from digital mammograms using a fully automated method [9].

Methods

Study design and participants

A genetic association meta-analysis was conducted within the Karolinska Mammography Project for Risk Prediction of Breast Cancer (Karma), a screening-based cohort of women attending one of four mammography units in the Swedish national screening program between 2011 and 2013 [10]. Participants responded to a web-based questionnaire, donated blood, and gave permission for storage of raw full-field digital mammograms. Genotyping data passing quality control metrics were available for two subcohorts of Karma participants with MD measurements, including 5827 women genotyped with the OncoArray and 4021 women genotyped with the Infinium iSelect genotyping array (Illumina, San Diego, CA, USA) of the Collaborative Oncological Gene-environment Study (iCOGS) (see details below). All women were free of cancer at the time of blood

sampling and had no history of breast enlargement, reduction, or surgery.

Volumetric MD assessment

MD in Karma was estimated from the mediolateral oblique view of baseline screening mammograms using Volpara™ version 1.5.0 (Matakina, Wellington, New Zealand), which is a fully automated method for volumetric MD estimation approved by the U.S. Food and Drug Administration. Volpara MD measures show good agreement with breast magnetic resonance imaging data [9] and have been validated as being predictive of breast cancer risk [11, 12]. Technical details of the Volpara software have been described in detail elsewhere [9]. In brief, the algorithm computes the thickness of dense tissue at each pixel using the X-ray attenuation of an entirely fatty region as an internal reference. The absolute dense volume (cm^3) is measured by integrating the dense thickness at each pixel over the whole mammogram, and the total breast volume (cm^3) is derived by multiplying the breast area by the recorded breast thickness, with an appropriate correction for the breast edge. From these measures, the absolute nondense volume (cm^3) and the percentage of the breast covered by dense tissue (%) can be obtained. The mean volumetric measurement of the left and right breast was taken for all analyses. Distributions of the three MD measures (percent dense volume, absolute dense volume, and absolute nondense volume) in the Karma-OncoArray and Karma-iCOGS genotyping cohorts are presented in Additional file 1: Figure S1.

Covariates

The following covariates were extracted from the Karma web-based questionnaire administered at baseline: age at study entry, menstrual and reproductive factors, and body mass index (BMI) as estimated from self-reported height and weight. Menopausal status was defined according to reported menstruation status, previous oophorectomy, and age. Women were considered postmenopausal if they had not menstruated during the past year, had a history of oophorectomy, or were above the age of 55 years.

Genotyping, quality control, and imputation

Whole-blood samples were genotyped using the OncoArray (see <https://support.illumina.com/downloads/infinium-oncoarray-500k-v1-0-product-files.html> for detailed information on array design), which covers more than 500,000 variants, or the iCOGS array (<http://ccge.medschl.cam.ac.uk/research/consortia/icogs/>), with over 200,000 variants. Both arrays use custom bead chips with markers of interest for breast and other cancers, fine-mapping of known susceptibility loci, and markers

associated with quantitative phenotypes that correlate with common cancers. The OncoArray is the successor of iCOGS and therefore denser, including additional variants of sequencing studies and a more informative backbone of approximately 260,000 SNPs that provides genome-wide coverage of most common variants. More details on both genotyping arrays can be found elsewhere [13, 14].

Samples were excluded from analysis for any of the following reasons: gender discordance according to array data, chromosomal anomalies (XXY or XO), extremely low or high heterozygosity (4.89 SD from the mean for the ethnicity), discordant duplicate pairs, first-degree relatives, and low call rates (*see* Additional file 1: Table S1 for number of individuals excluded per criterion). Standard SNP quality control was performed in PLINK version 1.9 [15], and SNPs with minor allele frequency (MAF) < 0.01 or deviation from Hardy-Weinberg equilibrium at $P < 1 \times 10^{-6}$ were excluded. To increase resolution and coverage, nongenotyped SNPs were imputed using the 1000 Genomes Project March 2012 release as the reference [16]. Data were imputed in a two-stage procedure using SHAPEIT to derive phased genotypes and IMPUTE version 2 to perform the imputation on the phased data [17]. Postimputation quality control was based on the IMPUTE info score, and SNPs with a score < 0.80 or MAF < 0.01 were excluded, resulting in a total of 8.5 million SNPs for analyses.

SNP association analyses

All three mammographic phenotypes (percent dense volume, absolute dense volume, and absolute nondense volume) were log-transformed to approximate the normal distribution (Additional file 1: Figure S2). SNP association analyses were performed in each genotyping cohort using linear regression and assuming an additive genetic model. Imputed SNPs were analyzed using a score test that employs allele dosages instead of hard genotype calls [18]. Population stratification was assessed using principal component analysis (PCA) in PLINK version 1.9 [15], and analyses were adjusted for age (years), BMI (kg/m^2), menopausal status (postmenopausal vs. premenopausal), and six cohort-specific PCA scores to account for population substructure. β -Coefficients of the two genotyping cohorts were meta-analyzed using a fixed effects model as implemented in METAL [19] with the Cochran's Q statistic used as a test for between-study heterogeneity. Regional association plots of identified variants were generated using LocusZoom [20] with the 400-kb region centered on the lead SNP.

Functional annotation and breast cancer association analyses of newly identified variants

We used several web tools for the functional annotation of the lead SNPs and their proxies ($r^2 > 0.80$). We

checked for potential regulatory functions using the HaploReg [21] and Regulome [22] databases, based on Encyclopedia of DNA Elements data [23] for the human mammary epithelial cell (HMEC) and mammary tumor cell (MCF-7) lines. We further searched the publicly available Genotype-Tissue Expression Project database (<http://www.gtexportal.org/>) for evidence of cis expression quantitative trait loci (eQTL) at each locus in mammary tissue samples ($n = 183$).

Associations between newly identified MD variants and breast cancer risk (overall and by estrogen receptor [ER] status) were checked using data from the Breast Cancer Association Consortium (BCAC), including 122,977 breast cancer cases and 105,974 control subjects (<http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/>) [14]. Information on ER status was available for 90,969 cases, of which 69,501 were ER-positive and 21,468 were ER-negative. We also verified associations with MD loci identified by previous GWAS [3–7], with the threshold of statistical significance set to $P = 0.05/13$ (number of loci) = 3.85×10^{-3} .

SNP-based heritability analyses

The heritability of a trait is defined as the proportion of phenotypic variance explained by genetic factors, including additive genetic effects, dominant effects, and epistasis. Narrow-sense heritability (h^2) refers to the variance component corresponding to additive genetic effects and can be estimated by exploring phenotypic similarities between relatives in family or twin studies.

Genome-wide complex trait analysis (GCTA) software [24, 25] was used to estimate the proportion of variance explained by additive effects of all SNPs. The interpretation of h^2_{SNP} estimated with GCTA is different from the h^2 obtained from traditional family-based studies, because the latter captures variance due to additive effects of all causal variants in the genome (including rare variants) and can be inflated if family resemblance is influenced by nonadditive genetic effects (dominance and epistasis or gene-gene interactions), shared environmental effects, and/or gene-environment interactions. Our GCTA analysis was conducted in the Karma-OncoArray cohort only, because the iCOGS array has insufficient coverage to obtain reliable genome-wide estimations of SNP-based heritability. First, pairwise genetic relationships between all individuals were calculated, followed by estimation of the additive genetic variance explained by all SNPs using restricted maximum likelihood analysis. We excluded one individual per pair whose estimated coefficient of relatedness was > 0.025 (which corresponds to third- or fourth-degree cousins), in order to prevent confounding by possible shared environmental effects and effects of causal variants that are not

tagged by the SNPs. This resulted in a study population of 4948 women for which h^2_{SNP} could be estimated for percent, absolute dense, and nondense MD. The covariates included were the same as those for the individual SNP analyses, and estimates of SNP-based heritabilities were compared with corresponding h^2 estimates previously reported in a study of siblings in Karma [26].

Results

Table 1 summarizes the characteristics of the Karma-OncoArray and Karma-iCOGS genotyping cohorts. Karma-OncoArray women were older and more often postmenopausal than women of the Karma-iCOGS cohort. Because of the older age distribution, mean MD levels were lower in the Karma-OncoArray cohort (see also Additional file 1: Figure S1 for the MD distributions per genotyping cohort). No major differences in BMI or family history of breast cancer were observed between the two cohorts.

Quantile-quantile plots of the genome-wide meta-analysis results for each MD measure are presented in Additional file 1: Figure S3. Overall, genomic inflation factors showed little or no evidence for inflation (λ for percent dense volume, absolute dense volume, and absolute nondense volume were 1.02, 1.03, and 1.02, respectively). Additional file 1: Figure S4 shows the Manhattan plots displaying the \log_{10} -transformed P values for each SNP. In total, we identified eight independent loci for any MD measure (*ZNF365*, *TAB2*, *HABP2*, *INHBB*, *AREG*, *LINC01483*, *MKLI*, and *8p11.23*) at $P < 5 \times 10^{-8}$ (Additional file 1: Table S2), three of which were novel (*HABP2*, *INHBB*, and *LINC01483*) and the remaining five

(*ZNF365*, *TAB2*, *AREG*, *8p11.23*, and *MKLI*) of which were previously reported to be associated with MD [3, 5, 7, 27] in the same directions as observed in the present analysis. We further replicated five of eight MD loci identified by previous GWAS (*1q12.21*, *ESR1*, *CCDC170/ESR1*, *IGF1*, and *SGSM3/MKLI*) beyond the loci reaching genome-wide significance in pooled analysis (Additional file 1: Table S3). The association between *PRDM6* and percent dense volume was marginally significant ($P < 0.05$), with direction of effect being consistent with that reported for percent dense area [5]. No significant associations of *TMEM184B* and *LSP1* with volumetric MD were observed. Altogether, the newly identified and established MD loci explained only small fractions of the total variance in percent dense (1.6%), absolute dense (2.8%), and absolute nondense (0.5%) volume.

The lead SNPs at the three newly identified MD loci (*HABP2*, *INHBB*, and *LINC01483*) are summarized in Table 2 with corresponding regional association plots in Fig. 1. SNP rs2089176 (10q25.3) associated with percent dense volume lies 120 kb upstream of its closest gene, *HABP2*. The two index SNPs associated with absolute dense volume span noncoding parts of the genome. SNP rs12468790 at 2q14.2 is located 11 kb upstream of the *INHBB* gene (Fig. 1) and rs9302903 at chromosome 17q24.3 falls within a long intergenic noncoding RNA gene (long intergenic non-protein coding RNA 1483 [*LINC01483*]) in a region flanking the mitogen-activated protein kinase 6 (*MAP2K6*) and the potassium voltage-gated channel subfamily J, member 16 (*KCNJ16*) genes. Of note, a variant (rs4849887) downstream of the *INHBB* gene but not in linkage disequilibrium (LD) with rs12468790 is an established breast cancer susceptibility variant [28, 29]. The association of rs12468790 with absolute dense volume remained significant in conditional analysis adjusting for rs4849887 ($P = 1.5 \times 10^{-8}$), supporting the presence of two independent signals at this locus.

For all newly identified loci, no predicted functional eQTL consequences were found in breast mammary tissue, but several variants in strong LD with rs2089176 (*HABP2*) mapped to a region defined by DNase and enhancer histone marks in HMECs and MCF-7 cells. In particular, one variant (rs1472751) was identified in a region binding CCCTC-binding factor (CTCF) in HMECs and MCF-7 cells, as well as GATA3 in MCF-7 cells. CTCF is a transcription factor that regulates a wide range of genes involved in growth, proliferation, differentiation, and apoptosis; GATA3 is the most abundant transcription factor in luminal epithelial cells and plays a key role in normal development of the mammary gland. Two variants in LD with rs9302903 (*LINC01483*) also mapped to DNase and enhancer histone marks in HMECs. No evidence of regulatory function was found for rs12468790 (*INHBB*) (Additional file 1: Table S4).

Table 1 Characteristics of the Karma study population

	Karma-OncoArray ($n = 5827$)	Karma-iCOGS ($n = 4021$)
Age, years, mean (SD)	60.2 (9.2)	53.6 (9.4)
Body mass index, kg/m ² , mean (SD)	25.5 (4.2)	25.3 (4.2)
Menopausal status, % (n)		
Premenopausal	22.8 (1327)	49.0 (1969)
Postmenopausal	77.2 (4500)	51.0 (2052)
Percent dense volume (%), median (IQR)	6.7 (5.2)	8.4 (6.5)
Dense volume, cm ³ , median (IQR)	53.4 (31.1)	60.3 (36.9)
Nondense volume, cm ³ , median (IQR)	744 (616)	676 (581)
Family history of breast cancer, % (n)	13.7 (769)	12.0 (482)

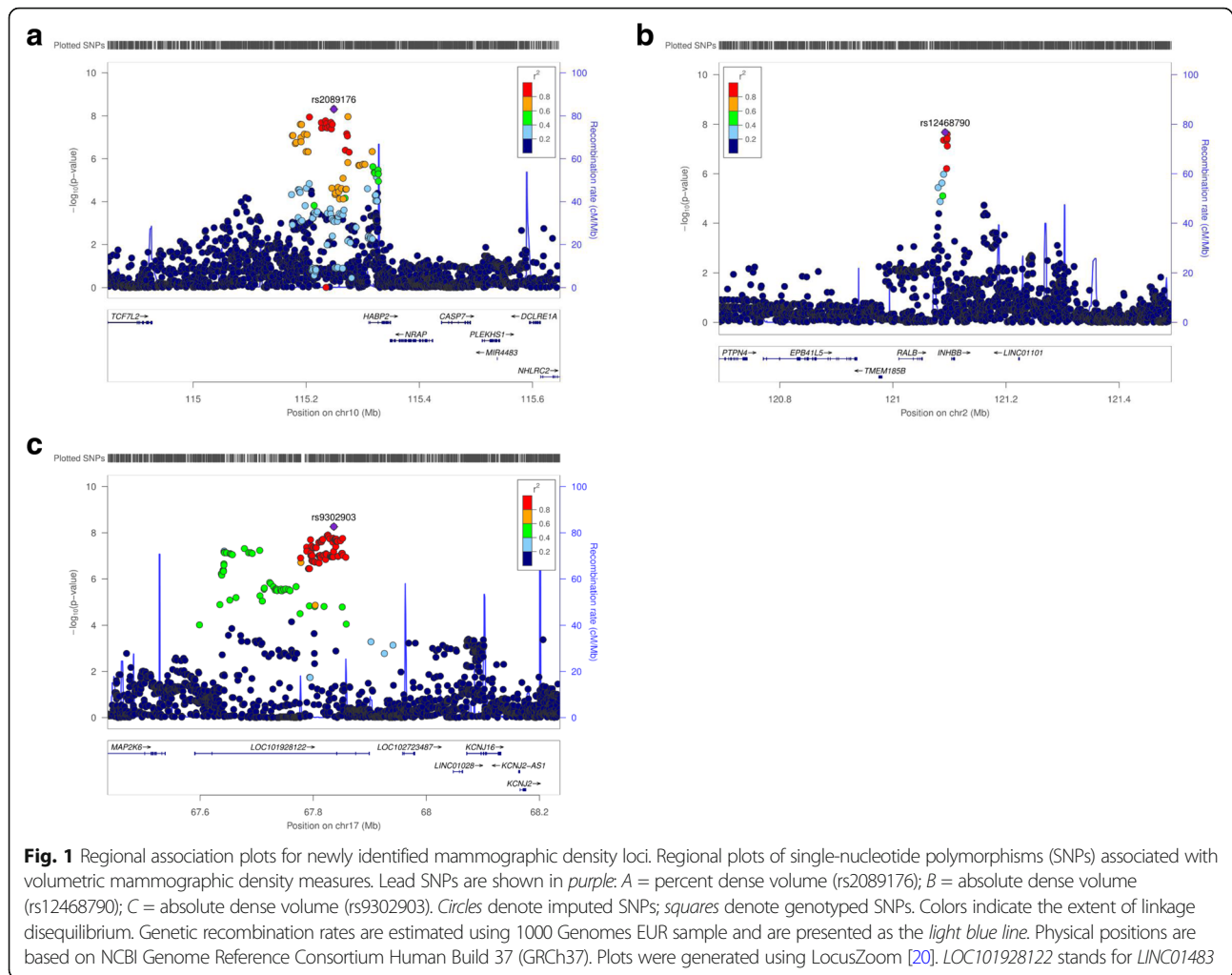
Karma Karolinska Mammography Project for Risk Prediction of Breast Cancer
Descriptive statistics of the Karma-OncoArray and Karma-iCOGS genotyping cohorts
Abbreviations: SD standard deviation, IQR interquartile range

Table 2 Novel loci associated with volumetric mammographic density

Locus	Base pair	Lead SNP	Gene	A1	A2	Karma-OncoArray (n = 5827)			Karma-iCOGS (n = 4021)			Karma pooled MD analysis (N = 9848)		
						MAF	β (SE)	P value	MAF	β (SE)	P value	β (SE)	P overall	P value for heterogeneity
Percent dense volume														
10q25.3	115,248,851	rs2089176	HABP2	A	G	0.41	-0.05 (0.01)	3.2 × 10 ⁻⁴	0.41	-0.08 (0.02)	1.4 × 10 ⁻⁶	-0.07 (0.01)	4.1 × 10 ⁻⁹	0.20
Absolute dense volume														
2q14.2	121,092,388	rs12468790	INHBB	A	G	0.39	0.08 (0.02)	2.4 × 10 ⁻⁵	0.36	0.08 (0.02)	2.3 × 10 ⁻⁴	0.08 (0.01)	2.1 × 10 ⁻⁸	0.84
17q24.3	67,836,371	rs9302903	LINC01483	T	C	0.08	-0.12 (0.03)	7.5 × 10 ⁻⁵	0.08	-0.17 (0.04)	1.3 × 10 ⁻⁵	-0.14 (0.02)	5.9 × 10 ⁻⁹	0.36

Abbreviations: iCOGS Illumina iSelect genotyping array of the Collaborative Oncological Gene-environment Study, Karma Karolinska Mammography Project for Risk Prediction of Breast Cancer, SNP Single-nucleotide polymorphism, A1 Major allele, A2 Minor allele, MAF Minor allele frequency

Gene refers to nearest gene. All lead SNPs were imputed: INFO scores Karma-OncoArray: rs2089176 (0.99), rs12468790 (0.96), rs9302903 (0.99)



Associations of the three MD loci with breast cancer risk are summarized in Table 3. rs2089176 (*HABP2*) and rs9302903 (*LINC01483*) were associated with breast cancer risk in BCAC meta-analysis data (rs2089176 OR [95% CI] per minor allele increase = 0.98 [0.97–0.99], $P = 0.01$; rs9302903 OR [95% CI] per minor allele increase = 0.96 [0.94–0.99], $P = 1.2 \times 10^{-3}$) in a direction consistent with that observed in the pooled MD analysis for percent dense volume [rs2089176 β [SE] per minor allele increase = -0.07 (0.01)] and absolute dense volume [rs9302903: β [SE] per minor allele increase = -0.14 (0.02)]. rs12468790 (*INHBB*) was also associated with breast cancer but in a direction opposite to its effect on MD. Each minor allele increase at rs12468790 resulted in a larger absolute dense volume (β [SE] = -0.03 [0.01]) but a lower odds of breast cancer (OR [95% CI] = 0.97 [0.96–0.98], $P = 4.8 \times 10^{-6}$). Stratified analyses by breast cancer subtype revealed that the association of rs12468790 (*INHBB*) with breast cancer was stronger for ER-negative (OR [95% CI] per minor allele increase = 0.93 [0.91–0.95], $P = 1.5 \times 10^{-9}$) than

ER-positive tumors (OR [95% CI] = 0.98 [0.96–0.99], $P = 3.8 \times 10^{-3}$).

Finally, we estimated the MD variance attributable to common genetic variants in the Karma-OncoArray cohort. SNP-based heritability (h^2_{SNP}) estimates for percent dense, absolute dense, and absolute nonvolume were 0.29 (SE = 0.07), 0.31 (SE = 0.07), and 0.25 (SE = 0.07) respectively, with evidence of a moderate genetic overlap between the absolute dense and nondense volume ($r_g = 0.45$, SE = 0.14). Given the narrow-sense h^2 estimates observed in previous sibling analyses in Karma ($h^2 = 0.63$ [SE = 0.06], 0.43 [SE = 0.06], and 0.61 [SE = 0.06] for percent, absolute dense, and absolute nondense volume, respectively [26]), the ratio of h^2_{SNP} to these narrow-sense h^2 estimates was substantially higher for the absolute dense volume (0.72) than for percent dense (0.46) and absolute nondense (0.41) volume.

Discussion

In this study, we identified three novel MD loci at genome-wide significance (percent dense volume

Table 3 Associations of newly identified mammographic density loci with breast cancer risk, overall and stratified by estrogen receptor status

Locus	Base pair	Lead SNP	Gene	A1	A2	Karma		BCAC						
						Pooled MD analysis		Breast cancer overall		Breast cancer ER–		Breast cancer ER+		
						β (SE)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	
Percent dense volume														
10q25.3	115,248,851	rs2089176	HABP2	A	G	–0.07 (0.01)	4.1×10^{-9}	0.36	0.98 (0.97–0.99)	0.01	1.00 (0.98–1.03)	0.70	0.97 (0.96–0.99)	5.3×10^{-4}
Absolute dense volume														
2q14.2	121,092,388	rs12468790	INHBB	A	G	0.08 (0.01)	2.1×10^{-8}	0.40	0.97 (0.96–0.98)	4.8×10^{-6}	0.93 (0.91–0.95)	1.5×10^{-9}	0.98 (0.96–0.99)	3.8×10^{-3}
17q24.3	67,836,371	rs9302903	LINC01483	T	C	–0.14 (0.02)	5.9×10^{-9}	0.09	0.96 (0.94–0.99)	1.1×10^{-3}	0.98 (0.94–1.02)	0.26	0.95 (0.93–0.98)	1.5×10^{-4}

Abbreviations: BCAC Breast Cancer Association Consortium, SNP Single-nucleotide polymorphism, A1 Major allele, A2 Minor allele, MAF Minor allele frequency, ER– Estrogen receptor-negative, ER+ Estrogen receptor-positive
MAF in BCAC as observed in genotyped control subjects (Europeans only). Gene refers to nearest gene

[*HABP2*] and absolute dense volume [*INHBB*, *LINC01483*]), of which two (*HABP2*, *LINC01483*) represent putative novel breast cancer susceptibility variants. We further demonstrate that at least 25% of the variance in volumetric MD is explained by common genetic variants, and that the ratio of SNP-based to narrow-sense heritability differs between the absolute dense and nondense volume.

All identified MD loci map to noncoding areas of the genome with nearby genes that have been linked to breast cancer etiology and/or mammary development. SNP rs2089176 upstream of the *HABP2* gene at 10q25.3 spans a region with CTCF and GATA3 transcription factor binding sites. The *HABP2* gene encodes for an extracellular serine protease that, upon binding to its ligand hyaluronic acid (HA), activates degradation of the extracellular matrix, including disruption of the endothelium, promoting tumor angiogenesis and cancer metastasis. In patients with breast cancer, a high HA content in malignant epithelial and stromal cells [30, 31] and elevated circulating HA levels [32, 33] have been associated with poor tumor differentiation, unfavorable prognostic features, and a reduction in MD. The protein encoded by *INHBB*, a subunit of both inhibin and activin (two glycoproteins belonging to the transforming growth factor- β superfamily), is critical in normal mammary development because its loss is accompanied by retarded ductal elongation and alveolar morphogenesis as well as failure of lactation [34]. Variants upstream of *INHBB* have previously been associated with bra cup size in a large European cohort of 16,000 women [35] but not with mammographic dense tissue specifically. Interpretation of the MD locus at 17q24.3 (*LINC01483*) is more speculative, because long intergenic noncoding RNAs are not well-characterized. Functional annotation revealed several correlated variants mapping to enhancer histone marks in HMECs, which may indicate a role of this locus in the expression of nearby genes (*MAP2K6* and *KCNJ16*), of which *MAP2K6* is known to play a role in the malignant transformation of breast epithelial cells [36, 37].

Finding genetic variants associated with both MD and breast cancer increases the understanding of biological pathways shared by both traits. Moreover, discovery of new MD loci in genome-wide approaches may provide information about putative novel susceptibility loci for breast cancer and/or its subtypes. Of the newly identified loci, two (*HABP2* and *LINC01483*) showed associations with MD that are consistent with the MD-breast cancer risk association, suggesting that part of their effect on breast cancer susceptibility is mediated through a directionally consistent change in MD. The association of rs12468790 (*INHBB*) with the absolute dense volume,

however, was not consistent with the MD-breast cancer risk association. Although the exact nature of this conflicting direction of effect is unknown, it may represent potential mediation by proxies closely related to MD that exert differential effects on breast cancer risk. *INHBB*, for instance, is highly expressed in adipose tissue [38] and has previously been linked to total breast size, a strong correlate of the absolute nondense volume [35, 39]. Albeit not reaching statistical significance, the *INHBB* variant was weakly associated with the absolute nondense volume in our study (rs12468790: β [SE] per minor allele increase = 0.03 [0.01], $P = 0.01$), and recent data indicate that dense and nondense adipose tissues are associated with breast cancer in opposite directions [40, 41]. The stronger association of this variant with ER-negative than ER-positive breast cancers may also point to alternative mechanisms linking *INHBB* to breast cancer than a causal path acting through mammographic dense tissue.

Apart from identifying novel MD loci, the present study replicates many loci identified by previous GWAS of area-based MD measures, except for *TMEM184B*, *LSPI*, 8p11.23, and 12q24 [4, 5]. Because our study concerns the volume of dense tissue in the breast rather than its projection, some inconsistency with previous GWAS is anticipated because volumetric and area-based measures reflect slightly different aspects of MD [7, 11, 26, 42]. Area-based and volumetric measures show good levels of agreement for percent MD ($r \approx 0.9$), but less so for absolute MD ($r \approx 0.5$) [11, 12]. Nevertheless, both percent and absolute MD measures show equivalent associations with breast cancer risk [11, 43, 44], regardless of measurement type. Lack of association in the Karma-iCOGS cohort, which included larger numbers of premenopausal women, may also reflect differential SNP effects over the life course. Previous GWAS were based largely on postmenopausal women, and some of the loci identified by these efforts and in the Karma-OncoArray cohort may not generalize to younger women. Differences in study design may further account for some of the differences observed. In contrast to previous GWAS, our study population did not include breast cancer cases with prediagnostic MD measurements. This reduces the likelihood of spurious associations due to confounding by breast cancer but may also have resulted in limited power to identify MD loci because of less extreme MD variation in cancer-free women.

Altogether, the newly identified and established MD loci explained only small fractions of the total variance in volumetric MD. More information about the genetic architecture of MD can be obtained by comparing SNP-based (h^2_{SNP}) to family-based (h^2) heritability estimates, with the missing heritability gap representing

the contribution of rarer genetic variants (to be discovered by whole-genome sequencing), gene-gene or gene-environment interactions, and/or possible inflation of family-based estimates when shared environmental effects are not specified in the model, such as is the case in sibling-based designs for h^2 estimation. Like other complex traits, the h^2_{SNP} estimates for percent dense and absolute nondense volume were approximately half of the h^2 estimates reported in our sibling study [26] (ratio h^2_{SNP} to $h^2 = 0.46$ and 0.41 , respectively). Interestingly, the heritability gap was much smaller for the absolute dense volume (ratio h^2_{SNP} to $h^2 = 0.76$). Because percent dense and absolute nondense volume are highly correlated with measures of adiposity [12, 26], this finding may suggest that sibling-based h^2 estimates of percent and absolute nondense volume are more likely to be confounded by shared environmental influences such as body fatness than sibling-based h^2 estimates of the absolute dense volume.

To our knowledge, this is the largest genetic association study of volumetric MD to date. All mammograms were derived from the same view and analyzed using fully automated software, reducing the likelihood of random measurement error. Genotyping data were complemented with imputed variants using 1000 Genomes Project data, resulting in high genome-wide coverage. We further reproduced associations with several established MD loci. Although a fully automated method such as Volpara has clear benefits in terms of a standardized and objective MD measurement, the underlying physics model tends to underestimate MD in very dense breasts [45–47]. Together with the narrower distribution of volumetric MD compared with area-based MD, this error could have resulted in reduced statistical power to identify novel MD loci and potential lack of association in the Karma-iCOGS cohort including women with more dense breasts. Furthermore, because our study population was restricted to cancer-free women, we were unable to address the role of volumetric MD in the mediation of SNP effects on breast cancer risk.

Conclusions

In summary, we report three novel MD loci at genome-wide significance, of which *HABP2* and *LINC01483* may represent putative new breast cancer susceptibility loci. We further demonstrate that 25% of the variance in volumetric MD is attributable to common genetic variation and that the ratio of SNP-based to narrow-sense heritability estimates varies between mammographic dense and nondense tissue components. Altogether, these findings provide more insight into the genetic basis of MD and mechanisms through which MD influences breast cancer risk.

Additional file

Additional file 1: Supplementary Tables and Figures. (DOCX 8763 kb)

Abbreviations

A1: Major allele; A2: Minor allele; BCAC: Breast Cancer Association Consortium; BMI: Body mass index; CTCF: CCCTC-binding factor; eQTL: Expression quantitative trait loci; ER: Estrogen receptor; GCTA: Genome-wide complex trait analysis; GWAS: Genome-wide association study; h^2 : Narrow-sense family-based heritability; h^2_{SNP} : Single-nucleotide polymorphism-based heritability; HA: Hyaluronic acid; HABP2: Hyaluronan-binding protein 2; HMEC: Human mammary epithelial cells; iCOGS: Illumina iSelect genotyping array of the Collaborative Oncological Gene-environment Study; INHBB: Inhibin beta B subunit; Karma: Karolinska Mammography Project for Risk Prediction of Breast Cancer; KCNJ16: Potassium voltage-gated channel subfamily J, member 16; LD: Linkage disequilibrium; LINC01483: Long intergenic non-protein coding RNA 1483; MAF: Minor allele frequency; MAP2K6: Mitogen-activated protein kinase 6; MCF-7: Michigan Cancer Foundation 7; MD: Mammographic density; PCA: Principal component analysis; SNP: Single-nucleotide polymorphism

Acknowledgements

The authors thank all participants in the Karma study, the study personnel for their devoted work during data collection, and Dr. Ralph Highnam and colleagues for their technical support with the Volpara software. The authors also thank the BCAC initiatives for sharing breast cancer association results for selected SNPs.

Funding

This work was financed by the Swedish Research Council (grant 2014- 2271), the Swedish Cancer Society (grant CAN 2016/684), and the Cancer Society in Stockholm (grant 141092). The Karma study is supported by the Märit and Hans Rausing Initiative Against Breast Cancer and the Cancer and Risk Prediction Center (CRiSP), a Linnaeus center (grant 70867902) financed by the Swedish Research Council. Genotyping of the OncoArray was principally funded by three sources: the PERSPECTIVE project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the *Ministère de l'Économie, de la Science et de l'Innovation du Québec* through Genome Québec, and the Quebec Breast Cancer Foundation; the National Cancer Institute Genetic Associations and Mechanisms in Oncology (GAME-ON) initiative and Discovery, Biology and Risk of Inherited Variants in Breast Cancer (DRIVE) projects (National Institutes of Health [NIH] grants U19 CA148065 and X01HG007492); and Cancer Research UK (C1287/A10118 and C1287/A16563). BCAC is funded by Cancer Research UK (C1287/A16563), the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS), and the European Union's Horizon 2020 Research and Innovation Programme under grant agreements 633784 (B-CAST) and 634935 (BRIDGES). Genotyping of the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175); Cancer Research UK (C1287/A10710), the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program; and the Ministry of Economic Development, Innovation and Export Trade of Quebec (grant PSR-SIIRI-701). Combination of the GWAS data was supported in part by NIH Cancer Post-Cancer GWAS initiative grant U19 CA 148065 (DRIVE, part of the GAME-ON initiative). The study sponsors had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Availability of data and materials

Subject to participants' consent and legal requirements, data can be made available upon request made to the primary department of the corresponding author.

Authors' contributions

JSB and KC conceived of and designed the study. JSB performed the statistical analyses. All authors contributed to the interpretation of the data. JSB drafted the manuscript with input from KC and KH. All authors critically reviewed the manuscript. All authors are accountable for the accuracy and integrity of the work. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All participants signed informed consent forms, and the ethical review board at Karolinska Institutet approved the study (2010/958-31/1).

Consent for publication

All authors approved the manuscript and consented to its publication.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels Väg 12A, 171 77 Stockholm, Sweden. ²Human Genetics, Genome Institute of Singapore, Singapore, Singapore.

Received: 20 September 2017 Accepted: 9 March 2018

Published online: 17 April 2018

References

- McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomark Prev.* 2006;15(6):1159–69.
- Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, et al. Mammographic breast density as an intermediate phenotype for breast cancer. *Lancet Oncol.* 2005;6(10):798–808.
- Lindstrom S, Vachon CM, Li J, Varghese J, Thompson D, Warren R, et al. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. *Nat Genet.* 2011;43(3):185–7.
- Stevens KN, Lindstrom S, Scott CG, Thompson D, Sellers TA, Wang X, et al. Identification of a novel percent mammographic density locus at 12q24. *Hum Mol Genet.* 2012;21(14):3299–305.
- Lindstrom S, Thompson DJ, Paterson AD, Li J, Gierach GL, Scott C, et al. Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk. *Nat Commun.* 2014;5:5303.
- Fernandez-Navarro P, Gonzalez-Neira A, Pita G, Diaz-Uriarte R, Tais Moreno L, Ederra M, et al. Genome wide association study identifies a novel putative mammographic density locus at 1q12-q21. *Int J Cancer.* 2015;136(10):2427–36.
- Brand JS, Li J, Humphreys K, Karlsson R, Eriksson M, Ivarsson E, et al. Identification of two novel mammographic density loci at 6Q25.1. *Breast Cancer Res.* 2015;17:75.
- Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. *Phys Med Biol.* 1994;39(10):1629–38.
- Highnam R, Brady M, Yaffe M, Karssemeijer N, Harvey J. Robust breast composition measurement - Volpara™. In: Martí J, Oliver A, Freixenet J, Martí R, editors. *Digital mammography: IWD 2010 (Lecture Notes in Computer Science series, vol. 6136)*. Berlin: Springer; 2010. p. 342–9.
- Gabrielsson M, Eriksson M, Hammarstrom M, Borgquist S, Leifland K, Czene K, et al. Cohort profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). *Int J Epidemiol.* 2017; <https://doi.org/10.1093/ije/dyw357>.
- Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, et al. Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. *Breast Cancer Res.* 2014;16(5):439.
- Brand JS, Czene K, Shepherd JA, Leifland K, Heddbom B, Sundbom A, et al. Automated measurement of volumetric mammographic density: a tool for widespread breast cancer risk assessment. *Cancer Epidemiol Biomark Prev.* 2014;23(9):1764–72.
- Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomark Prev.* 2017;26(1):126–35.
- Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92–4.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
- 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56–65.
- Howe B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955–9.
- Marchini J, Howe B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39(7):906–13.
- Waller CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190–1.
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26(18):2336–7.
- Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(Database issue):D930–4.
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790–7.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76–82.
- Yang J, Lee SH, Goddard ME, Visscher PM. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol.* 2013;1019:215–36.
- Brand JS, Humphreys K, Thompson DJ, Li J, Eriksson M, Hall P, et al. Volumetric mammographic density: heritability and association with breast cancer susceptibility loci. *J Natl Cancer Inst.* 2014;106(12):dju334.
- Stone J, Thompson DJ, Dos Santos SI, Scott C, Tamimi RM, Lindstrom S, et al. Novel associations between common breast cancer susceptibility variants and risk-predicting mammographic density measures. *Cancer Res.* 2015;75(12):2457–67.
- Milne RL, Kuchenbaecker KB, Michailidou K, Beesley J, Kar S, Lindstrom S, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet.* 2017;49(12):1767–78.
- Douglas JA, Roy-Gagnon MH, Zhou C, Mitchell BD, Shuldiner AR, Chan HP, et al. Mammographic breast density—evidence for genetic correlations with established breast cancer risk factors. *Cancer Epidemiol Biomark Prev.* 2008;17(12):3509–16.
- Auvinen P, Tammi R, Parkkinen J, Tammi M, Agren U, Johansson R, et al. Hyaluronan in peritumoral stroma and malignant cells associates with breast cancer spreading and predicts survival. *Am J Pathol.* 2000;156(2):529–36.
- Masarwah A, Tammi M, Sudah M, Sutela A, Oikari S, Kosma VM, et al. The reciprocal association between mammographic breast density, hyaluronan synthesis and patient outcome. *Breast Cancer Res Treat.* 2015;153(3):625–34.
- Delpech B, Chevallier B, Reinhardt N, Julien JP, Duval C, Maingonnat C, Bastit P, Asselain B. Serum hyaluronan (hyaluronic acid) in breast cancer patients. *Int J Cancer.* 1990;46(3):388–90.
- Peng C, Wallwiener M, Rudolph A, Cuk K, Eilber U, Celik M, et al. Plasma hyaluronic acid level as a prognostic and monitoring marker of metastatic breast cancer. *Int J Cancer.* 2016;138(10):2499–509.
- Robinson GW, Hennighausen L. Inhibitors and activators regulate mammary epithelial cell differentiation through mesenchymal-epithelial interactions. *Development.* 1997;124(14):2701–8.
- Eriksson N, Benton GM, Do CB, Kiefer AK, Mountain JL, Hinds DA, et al. Genetic variants associated with breast size also influence breast cancer risk. *BMC Med Genet.* 2012;13:53.
- Song H, Ki SH, Kim SG, Moon A. Activating transcription factor 2 mediates matrix metalloproteinase-2 transcriptional activation induced by p38 in breast epithelial cells. *Cancer Res.* 2006;66(21):10487–96.
- Kim ES, Jeong JB, Kim S, Lee KM, Ko E, Noh DY, et al. The G12 family proteins upregulate matrix metalloproteinase-2 via p53 leading to human breast cell invasion. *Breast Cancer Res Treat.* 2010;124(1):49–61.
- Sjoholm K, Palming J, Lystig TC, Jennische E, Woodruff TK, Carlsson B, et al. The expression of inhibin beta B is high in human adipocytes, reduced by weight loss, and correlates to factors implicated in metabolic disease. *Biochem Biophys Res Commun.* 2006;344(4):1308–14.
- Li J, Foo JN, Schoof N, Varghese JS, Fernandez-Navarro P, Gierach GL, et al. Large-scale genotyping identifies a new locus at 22q13.2 associated with female breast size. *J Med Genet.* 2013;50(10):666–73.
- Petterson A, Graff RE, Ursin G, Santos Silva ID, McCormack V, Baglietto L, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst.* 2014;106(5):dju078.

41. Pettersson A, Hankinson SE, Willett WC, Lagiou P, Trichopoulos D, Tamimi RM. Nondense mammographic area and risk of breast cancer. *Breast Cancer Res.* 2011;13(5):R100.
42. Lokate M, Kallenberg MG, Karssemeijer N, Van den Bosch MA, Peeters PH, Van Gils CH. Volumetric breast density from full-field digital mammograms and its association with breast cancer risk factors: a comparison with a threshold method. *Cancer Epidemiol Biomark Prev.* 2010;19(12):3096–105.
43. Cheddad A, Czene K, Eriksson M, Li J, Easton D, Hall P, et al. Area and volumetric density estimation in processed full-field digital mammograms for risk assessment of breast cancer. *PLoS One.* 2014;9(10):e110690.
44. Astley SM, Harkness EF, Sergeant JC, Warwick J, Stavrinou P, Warren R, et al. A comparison of five methods of measuring mammographic density: a case-control study. *Breast Cancer Res.* 2018;20:10.
45. Gubern-Merida A, Kallenberg M, Platel B, Mann RM, Marti R, Karssemeijer N. Volumetric breast density estimation from full-field digital mammograms: a validation study. *PLoS One.* 2014;9(1):e85952.
46. van Engeland S, Snoeren PR, Huisman H, Boetes C, Karssemeijer N. Volumetric breast density estimation from full-field digital mammograms. *IEEE Trans Med Imaging.* 2006;25(3):273–82.
47. Kallenberg MG, van Gils CH, Lokate M, den Heeten GJ, Karssemeijer N. Effect of compression paddle tilt correction on volumetric breast density estimation. *Phys Med Biol.* 2012;57(16):5155–68.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

