

**Novel Associations between Common Breast Cancer Susceptibility Variants and Risk-
Predicting Mammographic Density Measures**

Running Title: Breast Cancer Susceptibility Loci and Mammographic Density

Jennifer Stone¹, Deborah J. Thompson^{2,3}, Isabel dos-Santos-Silva⁴, Christopher Scott⁵,
Rulla M. Tamimi⁶, Sara Lindstrom⁷, Peter Kraft⁸, Aditi Hazra⁹, Jingmei Li¹⁰, Louise Eriksson¹¹,
Kamila Czene¹¹, Per Hall¹¹, Matt Jensen⁵, Julie Cunningham¹², Janet E. Olson¹³,
Kristen Purrington¹⁴, Fergus J. Couch^{12,13}, Judith Brown³, Jean Leyland³, Ruth M. L. Warren¹⁵,
Robert N. Luben³, Kay-Tee Khaw², Paula Smith¹⁶, Nicholas J. Wareham¹⁷, Sebastian M. Jud¹⁸,
Katharina Heusinger¹⁸, Matthias W. Beckmann¹⁸, Julie A. Douglas¹⁹, Kaanan P. Shah¹⁹,
Heang-Ping Chan²⁰, Mark A. Helvie²⁰, Loic Le Marchand²¹, Laurence N. Kolonel²¹,
Christy Woolcott³⁸, Gertraud Maskarinec²¹, Christopher Haiman²², Graham G. Giles²³, Laura
Baglietto²³, Kavitha Krishnan²⁴, Melissa C. Southey²⁵, Carmel Apicella²⁶,
Irene L. Andrulis²⁷, Julia A. Knight²⁸, Giske Ursin²⁹, Grethe I. Grenaker Alnaes³⁰, Vessela N.
Kristensen³⁰, Anne-Lise Borresen-Dale³⁰, Inger Torhild Gram³¹, Manjeet K. Bolla³², Qin
Wang³¹, Kyriaki Michailidou³², Joe Dennis³², Jacques Simard³³,
Paul Paroah³⁴, Alison M. Dunning³⁵, Douglas F. Easton³⁶, Peter A. Fasching^{18,37}, V. Shane
Pankratz⁵, John Hopper²⁶, Celine M. Vachon¹³

¹Centre for Genetic Origins of Health and Disease, University of Western Australia; ²MRC
Centre for Nutritional Epidemiology in Cancer Prevention and Survival (CNC), University of
Cambridge, Cambridge, UK; ³Centre for Cancer Genetic Epidemiology, Department of Public
Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts

Causeway, Cambridge, CB1 8RN, UK; ⁴Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, UK; ⁵Department of Health Sciences Research, Division of Biostatistics Mayo Clinic College of Medicine; ⁶Channing Laboratory, Brigham and Women's Hospital, Boston, MA 02115, USA, and Department of Epidemiology, Harvard School of Public Health, Boston, MA, 02115; ⁷Program in Genetic Epidemiology and Statistical Genetics, Harvard School Of Public Health, Boston, MA, USA and Department of Epidemiology, Harvard School Of Public Health, Boston, MA, USA; ⁸Program in Molecular and Genetic Epidemiology and Department of Epidemiology and Department of Biostatistics, Harvard School Of Public Health, Boston, MA; ⁹Program in Molecular and Genetic Epidemiology and Department of Epidemiology, Harvard School of Public Health, Boston, MA and Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; ¹⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden, and Human Genetics, Genome Institute of Singapore, Singapore; ¹¹Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; ¹²Division of Experimental Pathology, Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, MN; ¹³Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, Rochester, MN 55905, USA; ¹⁴Wayne State University School of Medicine and Karmanos Cancer Institute, Detroit, MI; ¹⁵Department of Radiology, University of Cambridge, Addenbrooke's NHS Foundation Trust Cambridge, UK; ¹⁶Department of Public Health and Primary Care and Department of Psychiatry, University of Cambridge, Cambridge, UK; ¹⁷MRC Epidemiology Unit, University of Cambridge, Cambridge, UK; ¹⁸University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive

Cancer Center Erlangen-Nuremberg, Germany;¹⁹Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA;²⁰Department of Radiology, University of Michigan Medical School, Ann Arbor, MI, USA;²¹University of Hawaii Cancer Center, Honolulu, HI 96813, USA;²²Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA;²³Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia, and Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Australia;²⁴Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia;²⁵Department of Pathology, University of Melbourne, Melbourne, Australia;²⁶Centre for Epidemiology and Biostatistics Melbourne School of Population and Global Health, The University of Melbourne, Melbourne;²⁷Center for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada, and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada;²⁸Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada, and Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada;²⁹Institute of Basic Medical Sciences, University of Oslo, Norway & Department of Preventive Medicine, University of Southern California, CA, USA;³⁰Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Montebello 0310, Oslo, Norway;³¹Institute of Community Medicine, Faculty of Health Sciences, University of Tromsø, Norway;³²Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, CB1 8RN, UK;³³Centre Hospitalier Universitaire de Québec Research Center and Laval University, QC, G1V 4G2, Canada;³⁴Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts

Causeway, Cambridge, CB1 8RN, UK; ³⁵Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, CB1 8RN, UK; ³⁶Centre for Cancer Genetic Epidemiology, Department of Public Health, Primary Care and Oncology, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, CB1 8RN, UK; ³⁷University of California at Los Angeles, Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, USA; ³⁸Department of Obstetrics and Gynecology, IWK Health Centre, Halifax, NS B3K 6R8, Canada.

Corresponding Author

Celine M. Vachon, PhD; Mayo Clinic, 200 First Street SW, Charlton Building 6-239;
Rochester, MN 55905; Telephone: 507-284-9977 Fax: 507-284-1516;
E-mail: vachon.celine@mayo.edu

Grant Support

ABCFS: The Australian Breast Cancer Family Registry (ABCFR; 1992-1995) was supported by the Australian NHMRC, the New South Wales Cancer Council, and the Victorian Health Promotion Foundation (Australia), and by grant UM1CA164920 from the USA National Cancer Institute. The Genetic Epidemiology Laboratory at the University of Melbourne has also received generous support from Mr B. Hovey and Dr and Mrs R.W. Brown to whom we are most grateful. 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 Breast

Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR.

BBCC: This study was funded in part by the ELAN-Program of the University Hospital Erlangen; Katharina Heusinger was funded by the ELAN program of the University Hospital Erlangen. BBCC was supported in part by the ELAN program of the Medical Faculty, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg.

EPIC-Norfolk: This study was funded by research programme grant funding from Cancer Research UK and the Medical Research Council with additional support from the Stroke Association, British Heart Foundation, Department of Health, Research into Ageing and Academy of Medical Sciences.

MCBCS: This study was supported by Public Health Service Grants P50 CA 116201, R01 CA 128931, R01 CA 128931-S01, R01 CA 122340, CCSG P30 CA15083, from the National Cancer Institute, National Institutes of Health, and Department of Health and Human Services.

MCCS: Melissa C. Southey is a National Health and Medical Research Council Senior Research Fellow and a Victorian Breast Cancer Research Consortium Group Leader. The study was supported by the Cancer Council of Victoria and by the Victorian Breast Cancer Research Consortium.

MEC: National Cancer Institute: R37CA054281, R01CA063464, R01CA085265, R25CA090956, R01CA132839.

MMHS: This work was supported by grants from the National Cancer Institute, National Institutes of Health, and Department of Health and Human Services. (R01 CA128931, R01 CA 128931-S01, R01 CA97396, P50 CA116201, and Cancer Center Support Grant P30 CA15083).

NBCS: This study has been supported with grants from Norwegian Research Council (#183621/S10 and #175240/S10), The Norwegian Cancer Society (PK80108002, PK60287003), and The Radium Hospital Foundation as well as S-02036 from South Eastern Norway Regional Health Authority.

NHS: This study was supported by Public Health Service Grants CA131332, CA087969, CA089393, CA049449, CA98233, CA128931, CA 116201, CA 122340 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

OOA study was supported by CA122822 and X01 HG005954 from the NIH; Breast Cancer Research Fund; Elizabeth C. Crosby Research Award, Gladys E. Davis Endowed Fund, and the Office of the Vice President for Research at the University of Michigan. Genotyping services for the OOA study were provided by the Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096.

OFBCR: This work was supported by grant UM1 CA164920 from the USA National Cancer Institute. 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 Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR.

SASBAC: The SASBAC study was supported by Märit and Hans Rausing's Initiative against Breast Cancer, National Institutes of Health, Susan Komen Foundation and Agency for Science, Technology and Research of Singapore (A*STAR).

SIBS: SIBS was supported by program grant C1287/A10118 and project grants from Cancer Research UK (grant numbers C1287/8459).

COGS grant: Collaborative Oncological Gene-environment Study (COGS) that enabled the genotyping for this study. Funding for the BCAC component is provided by grants from the EU FP7 programme (COGS) and from Cancer Research UK. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

Conflicts of Interest

None of the authors have a conflict of interest.

Key Words: mammographic density; genetic variation; single nucleotide polymorphisms (SNPs); breast cancer; risk factors

Word Count: 3904

Number of Tables: 3

Number of Figures: 2

Abstract

Mammographic density measures adjusted for age and body mass index (BMI) are heritable predictors of breast cancer risk but few mammographic density-associated genetic variants have been identified. Using data for 10,727 women from two international consortia, we estimated associations between 77 common breast cancer susceptibility variants and absolute dense area, percent dense area and absolute non-dense area adjusted for study, age and BMI using mixed linear modeling. We found strong support for established associations between rs10995190 (in the region of *ZNF365*), rs2046210 (*ESR1*) and rs3817198 (*LSP1*) and adjusted absolute and percent dense areas (all $p < 10^{-5}$). Of 41 recently discovered breast cancer susceptibility variants, associations were found between rs1432679 (*EBF1*), rs17817449 (*MIR1972-2: FTO*), rs12710696 (2p24.1), and rs3757318 (*ESR1*) and adjusted absolute and percent dense areas, respectively. There were associations between rs6001930 (*MKL1*) and both adjusted absolute dense and non-dense areas, and between rs17356907 (*NTN4*) and adjusted absolute non-dense area. Trends in all but two associations were consistent with those for breast cancer risk. Results suggested that 18% of breast cancer susceptibility variants were associated with at least one mammographic density measure. Genetic variants at multiple loci were associated with both breast cancer risk and the mammographic density measures. Further understanding of the underlying mechanisms at these loci could help identify etiological pathways implicated in how mammographic density predicts breast cancer risk.

Precis: Findings significantly extend evidence of shared genetic determinants between breast cancer risk and mammographic density metrics, likely representing shared etiological pathways.

Introduction

Mammographic density refers to the white or light areas on a mammogram, which are thought to reflect differing amounts of epithelial and stromal tissue within the breast, as distinct from radiographically lucent fatty tissue. For women of the same age and body mass index (BMI), those with more extensive amounts of either absolute or percent dense area are more likely to develop breast cancer (1). The underlying biological processes are not clear.

Twin and family studies have shown that a substantial variation in the mammographic density measures could be due to genetic factors (2-4). Moreover, these heritable mammographic density measures are thought to explain about 10-20% of the association of family history with breast cancer risk (5, 6).

Finding genetic variants that are associated with both breast cancer risk and the mammographic density measures that predict breast cancer has the potential to reveal underlying biological pathways that explain the associations between those mammographic measures and cancer, resulting in a better understanding of the etiology of breast cancer. The use of large scale genotyping projects to discover common genetic variants (single nucleotide polymorphisms, or SNPs) associated with breast cancer risk has opened up the possibility of achieving this. The international DENSNP consortium previously studied the associations of 15 independent breast cancer susceptibility variants with age- and BMI-adjusted mammographic density measures for 17,000 women. This confirmed prior associations found between the variant rs381798 (in the region of *LSP1*) (7, 8) and adjusted absolute and percent density and provided evidence for an association between rs10483813 (in the region of *RAD51L1*) and adjusted percent dense area (9). Two genome-wide

association studies (GWAS) conducted by the Markers Of DEntity (MODE) consortium found that there was an association between rs10995190 (in the *ZNF365* locus), independently shown to be associated with breast cancer risk (10), and adjusted percent dense area, and weaker evidence for associations with the variants rs2046210 (in the region of *ESR1*) and rs3817198 (see above) (11). More recently, we identified novel loci associated with dense area (rs10034692 from *AREG*, rs703556 from *IGF1*, rs7289126 from *TMEM184B*, rs17001868 from *SGSME/MKL1*), non-dense area (rs7816345 from *8p11.23*), and percent density (rs186749 from *PRDM6*, rs7816345 from *8p11.23* and rs7289126 from *TMEM184B*) (11). Furthermore, using a GWAS of both breast cancer and mammographic density, MODE investigators found that adjusted percent dense area and breast cancer risk have a shared genetic basis that is mediated by, at least in theory, a large number of common variants (12).

A further 41 independent breast cancer susceptibility common variants have been discovered by a study of 45,290 cases and 41,880 controls using a custom genotyping array designed in part by the Breast Cancer Association Consortium (BCAC) (13). Of these new variants, a recent report from several co-authors found novel associations between breast cancer SNPs in 6q25: rs9485372 (*TAB2*) and rs9383938 (*ESR1*) with a volumetric measure of mammographic density in approximately 5000 Swedish women (14). They also found novel associations between breast cancer SNPs rs6001930 (*MKL1*) and rs17356907 (*NTN4*) with absolute non-dense volume. Here, we provide the largest and most comprehensive report to date of the associations between the current total of 77 known breast cancer susceptibility SNPs and three area-based mammographic density measures using data from over 10,000 women participating in the DENSNPs and MODE consortia.

Methods

Subjects

Genotypes, mammographic density measures and information on conventional breast cancer risk factors were available for 10,727 self-reported women of European Ancestry from 13 studies described previously (4, 9, 11, 15). A summary of study design, sample sizes, mammographic and genotyping characteristics is given in Supplementary Table 1. Each study obtained informed consent and had relevant ethics and institutional approvals. Only anonymised data were used for analyses.

Mammographic density measures

All mammographic density measurements were performed on digitized analogue films taken prior to diagnosis using either the Cumulus (16), Madena (17), or MDEST (18) programs. All approaches apply a thresholding technique to measure total area of the breast and absolute dense area, from which percent dense area and absolute non-dense area are derived. Absolute dense and non-dense area values were converted to cm^2 according to the pixel size used in the digitization. All measurements were conducted by observers blind to genotype, case status (if applicable) and breast cancer risk factor data. For cases, mammograms prior to diagnosis were used or, when this was not possible, those from the contralateral breast taken at the time of diagnosis (Table 1).

The mammographic density readings were performed on both craniocaudal (CC) and mediolateral oblique (MLO) views but these have been consistently shown to have high correlation (range of 0.87-0.90) (19).

Genotyping

The 77 currently known independent breast cancer susceptibility SNPs were genotyped for the 13 studies either as part of a GWAS (11, 15) or by genotyping of a custom Illumina iSelect genotyping array comprising 211,155 SNPs (described in Michailidou *et al* 2013 (13) (Table 1). Quality control was conducted at the study level; for all SNPs in these analyses their call rates were >95%. Five SNPs (from 3 studies) with Hardy–Weinberg equilibrium P values <0.001 were excluded.

Statistical methods

Distributions of covariates summarized by frequency and percentages are summarized breast cancer status (affected/unaffected). Primary analyses used individual level data and included a fixed study effect to adjust for potential differences due to study. Analyses were conducted using the square root of the density measures as the outcome variables, and examination of the distributions of the residuals after adjustment for age and BMI showed an approximately normal distribution.

Primary analyses were conducted using fixed effects ordinary linear regression adjusting for age (continuous), 1/BMI, and study. Analyses considered SNP associations as additive by defining an ordinal covariate as the number of copies of the minor allele (0, 1 or 2) producing per-allele estimates that are reported as beta (β) and standard error (SE). (For imputed genotypes from the two GWAS studies, the imputed allelic dosage values were used). Secondary analyses were performed to evaluate potential confounding with other covariates such as case-control status, menopausal status (pre- and perimenopausal

combined vs. postmenopausal), and postmenopausal hormone use (ever vs. never use). To measure the extent to which the mammographic measures mediated the SNP associations with breast cancer risk we estimated the proportion of change in the regression coefficient for each SNP after adjustment for breast cancer status and calculated 95% confidence intervals based on methods described by Lin *et al* (20).

We performed a series of analyses to test the robustness of the association between mammographic density measures and the 77 SNPs. First we performed an overall test of whether there was no association between any of the variants and a given mammographic measure by testing whether the distribution of the 77 P-values deviated from the uniform distribution on the interval 0, 1. Fisher's exact test of uniformity tests the sum of the $-2 \ln P_i$ across all loci where P_i is the P value for the i^{th} variant, against χ^2 distribution with $2n$ degrees of freedom, where n is the number of independent variants (21). Second, to try to determine the "best" model fit (i.e. the set of independent SNPs which give the best-fitting model for adjusted breast density) we used Lasso (least absolute shrinkage and selection operator) regression, a method which combines estimation and model selection which limits overestimation of associations when there are a large number of covariates (22). The final model was chosen by the minimum Schwarz Bayesian Information Criterion (SBC), which combines goodness of fit with a penalty based on the number of parameters in the model. Finally, we tried to quantify whether there was information in the other variants that did not reach our p-value threshold (see below for details) but which could help further explain some of the missing heritability. For each mammographic density measure we removed the most significant variants ($p < 0.00065$, selected by $0.05/77$) associated with that measure and tested whether the distribution of the remaining p-values was different from

zero. The least informative variant was removed sequentially until there was no evidence to reject the null hypothesis.

Analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC). Two sided p-values were calculated. We used a conservative threshold of $0.05/77=0.00065$ to define statistical significance, while presenting the results for all tested variants.

Results

Table 2 shows summary characteristics for each study. The majority of women were older than 60 years, more than 80% were postmenopausal, 55% had BMI ≥ 25 kg/m², and 35% were breast cancer cases.

Percent and absolute dense area were negatively associated with age, BMI, parity and postmenopausal status and positively associated with postmenopausal hormone therapy use (Supplementary Table 1). Conversely, absolute non-dense area was positively associated with age, BMI and parity and negatively associated with hormone therapy use. All of the above associations were similar in direction and magnitude for cases and controls (data not shown). None of the density measures were statistically significantly different by mammogram view (Supplementary Table 1).

Of the 77 variants, nine were associated with at least one adjusted mammographic density measure, using the threshold of 0.00065 (Table 3, results for all SNPs in Supplementary Table 1). Figure 1 is a forest plot of all 77 breast cancer susceptibility variants sorted by magnitude of association with breast cancer risk in these studies; the nine variants are highlighted in bold. The findings confirm previously identified associations with both adjusted percent and absolute dense areas for rs10995190 in the *ZNF365* gene ($\beta=0.16$, SE=0.028, $p=8.5 \times 10^{-9}$ and $\beta=0.25$, SE=0.038, $p=4.7 \times 10^{-11}$, respectively), rs2046210 in

the region of *ESR1* ($\beta=0.098$, $SE=0.021$, $p=2.4 \times 10^{-6}$ and $\beta=0.14$, $SE=0.029$, $p=1.7 \times 10^{-6}$, respectively) and rs3817198 in the region of *LSP1* ($\beta=0.087$, $SE=0.021$, $p=4.4 \times 10^{-5}$ and $\beta=0.16$, $SE=0.029$, $p=1.3 \times 10^{-7}$, respectively). None of these three variants showed evidence of association with adjusted non-dense area (Table 3). There were marginal associations between two independent variants ($r^2=0.003$) in the region of *RAD51L1*; rs999737 ($p=0.003$ and $p=0.01$) with both adjusted percent and absolute dense area (reported in our previous DENSNP study (9)) and rs2588809 ($p=0.002$, $p=0.04$, and $p=0.02$) with adjusted percent dense area, dense area and non-dense area respectively (Supplementary Table 2).

Of the 41 recently identified breast cancer loci, we found evidence of novel associations between at least one of the adjusted density measures and six variants (Table 3). The minor G allele of rs1432679 (*EBF1*) was positively associated with adjusted dense area and negatively associated with adjusted non-dense area, and hence was positively associated with adjusted percent density ($\beta=0.087$, $SE=0.020$, $p=1.1 \times 10^{-5}$). The minor G allele of rs6001930 in the region of *MKL1* was negatively associated with both adjusted absolute dense and non-dense areas ($\beta=-0.18$, $SE=0.044$, $p=3.2 \times 10^{-5}$ and $\beta=-0.23$, $SE=0.048$, $p=1.7 \times 10^{-6}$, respectively), but was not associated with adjusted percent density ($p=0.04$). The A allele of rs17356907 in the region of *NTN4* was negatively associated with adjusted non-dense area ($\beta=-0.12$, $SE=0.033$, $p=2.4 \times 10^{-4}$), but not with adjusted dense area or percent density. The A allele of rs3757318 (close to *ESR1*) was positively associated with adjusted dense area ($\beta=0.19$, $SE=0.054$, $p=4.6 \times 10^{-4}$), but not with either of the other density phenotypes. Both rs17817449 (*MIR1972-2:FTO*) and rs12710696 (2p24.1) were negatively associated with adjusted percent and absolute dense area. Although sample sizes were substantially reduced ($n < 7000$), these associations with were similar when analyses were

restricted to images from controls only, CC mammogram views, and mammograms within a year of covariate information (data not shown).

Further adjustment for case-control status showed evidence that percent dense area and dense area mediated the associations of rs10995190 (*ZNF365*), rs2046210 (*ESR1*), rs1432679 (*EBF1*), and rs3817198 (*LSP1*) with breast cancer risk (Supplementary Table 3). There was also evidence that dense area mediated the association of rs3757318 (*ESR1*) and breast cancer, and non-dense area mediated the association of rs1432679 (*EBF1*) and rs6001930 (*MKL1*) with breast cancer. These estimates ranged from 4% to 18% of the SNP and breast cancer association being explained by density phenotypes (Supplementary Table 3). However, adjustment for other additional covariates did not substantially influence the regression estimates (data not shown). The between-study test of heterogeneity p-value was >0.05 for all the variants in Table 3, except for the association between rs2046210 (*ESR1*) and adjusted dense area ($p=0.03$).

When taking a global, as distinct from individual SNP, view we found that of the 77 variants examined, the nominal p-value was <0.05 for 20 associations with adjusted dense area, 18 associations with adjusted percent dense area, and 10 associations with adjusted non-dense area comprising in total 25 separate variants (Supplementary Table 2). For any one density measure, by chance alone we would have expected 3.9 (95% CI 1-7) associations to be nominally significant at $p=0.05$. The distributions of the 77 P-values for each of the mammographic measures were not consistent with the uniform distribution ($p_{\text{uniform}} < 2 \times 10^{-6}$ for each density measure), suggesting the existence of true associations between at least some breast cancer susceptibility variants and the mammographic density measures that predict breast cancer. From Lasso regression, the 'best fitting' model for adjusted percent dense area included the bolded variants in Table 3 plus three others

including rs12710696 (2p24.1), rs4808801 (*SSBP4:ISYNA1:ELL*) and rs12422552 (12p13.1) which when combined explained 1.2% of the variation in the adjusted percent dense area trait. Similarly, the best fitting model from Lasso regression of adjusted dense area included the six bolded variants in Table 3 and rs1432679 (*EBF1*), rs17817449 (*MIR1972-2:FTO*) and, like percent dense area, rs4808801 (*SSBP4:ISYNA1:ELL*) when combined explained 1.4% of the variation in adjusted dense area. Finally, the best fitting model for adjusted non-dense area included only the three bolded variants in Table 3 which when combined explained 0.4% of the variation.

When we removed the variants in the best fitting Lasso models for each phenotype above noted above, the distribution of the remaining p-values still deviated from the null hypothesis of no association between the genetic variants for both percent and absolute dense area ($p=0.00005$ for percent density, $p=0.00006$ for dense area) but not non-dense area ($p=0.01$). Sequentially removing the variants most strongly associated with the mammographic density measure until the test no longer found deviation from the null using a p-value threshold of 0.00065, we found evidence of one more SNP associated with each percent dense and absolute dense area (rs999737 and rs6678914, respectively).

We also compared the QQ plots before and after exclusion of the top 14 breast cancer variants (9 identified via ordinary linear regression and 5 others identified via LASSO regression) most strongly associated with the mammographic density measures (see Figure 2). Based on the analysis described above and the lack of departure from the 45 degree line once the top 14 variants have been removed (indicative of a probable threshold of commonly shared variants), we estimate that there is an approximate 18% overlap between breast cancer- and mammographic density- associated variants.

Discussion

This is the largest and most comprehensive study to date of the associations between breast cancer risk-predicting mammographic density measures and the 77 independent established breast cancer susceptibility common variants, 41 of which were recently identified by large-scale genotyping (13). In addition to previously reported associations between common breast cancer susceptibility SNPs in the regions of *ZNF365*, *ESR1* and *LSP1* and age- and BMI-adjusted absolute and percent dense area, we found strong evidence of novel associations between SNPs with all three adjusted mammographic density measures (in the region of *EBF1*); with adjusted dense area or percent dense area (*MIR1972-2: FTO*, *2p24.1*, and another in the region of *ESR1* independent of the initially reported SNP, rs2046210); adjusted dense and non-dense area (*MKL1*) and non-dense area only (*NTN4*). The directions of these associations were consistent with that of their associations with breast cancer risk, with the exception of *MKL1* and 2p24.1 which were both negatively associated with dense area (and percent dense area) but are reportedly positively associated with breast cancer risk. Further, these mammographic measures show evidence for mediating the association of several of these SNPs and breast cancer risk.

These findings are consistent with those recently reported by our co-authors Swedish study (14). Despite differences in phenotypes (area vs volume), both studies independently showed novel associations between absolute measures of dense tissue with rs2046210 (*ESR1*) and between absolute measures of non-dense tissue with rs17356907 (*NTN4*). Both studies also reported strong negative associations between and absolute measures of dense and non-dense tissue with rs6001930 (*MKL1*). The other novel

association reported in Brand et al (14) between percent dense volume and rs9485372 (*TAB2*), a variant associated with breast cancer risk in Asian women, was not investigated in this study. The Swedish study did not replicate the previously reported association with rs3817198 (*LSP1*) nor our novel associations with rs12710696 (2p24.1) and rs17817449 (*MIR1972-2: FTO*), *underscoring the differences in the volumetric and area phenotypes*. Of note, both volumetric and area-based density measures have been shown associated with breast cancer risk, with similar magnitude of association (23).

Whilst the standard approach using linear regression identified nine variants associated with mammographic density, the non-uniform distributions of the remaining p-values suggest that there are additional genetic variants associated with both breast cancer risk and the mammographic density measures that predict risk. In total, there is evidence of at least 14 breast cancer susceptibility variants (18%) associated with at least one mammographic density measure; approximately 10%, 12% and 4% of the breast cancer susceptibility SNPs were associated with percent dense area, dense area and non-dense area, respectively. Our estimate of 18% is consistent with empirical estimates that the percentage of overlap between genetic determinants of breast cancer and the risk-predicting mammographic density measures is 14% (95% CI: 4-39%) (5, 12).

The nine density-associated variants identified here (using the standard approach) account for only a small proportion of the between-woman variation in the three risk-predicting mammographic density phenotypes (<1.5% for each), but the contribution of the true causal variants could be larger. Also, it has been estimated that there are more than 1000 loci involved with breast cancer susceptibility (13) and therefore it is possible that a considerable subset of these will also be associated with the mammographic density measures. Importantly, several of the SNP associations with breast cancer appear to be

mediated by the mammographic density phenotypes. In fact, 15-20% of the associations of variants at *ESR1* and *EBF1* with breast cancer were mediated by percent dense area or dense area. Understanding which susceptibility loci exert their influence on breast cancer risk partially through mammographic density measures could be important for identifying subgroups of women who are at a high “genetic” risk for both breast cancer and mammographic density. There is increasing demand for the evidence-base to support stratified breast screening programs instead of the “one-size-fits-all” approaches that are currently recommended in most countries. Discriminating between genetic risk due to mammographic density and/or breast cancer risk could identify which women may have a greater benefit from density reduction strategies and/or additional breast screening measures.

The biological reasons why mammographic density measures predict breast cancer risk are not understood. There is evidence to suggest that extensive mammographic density is causally related to breast cancer rather than a simple correlate of its determinants (24). Breast cancer arises from epithelial cells lining the ducts or lobules of the breast and mammographic density might represent areas of the breast in which there are higher rates of epithelial proliferation, which are likely to increase the risk of somatic mutations, epigenetic alterations, and carcinogenesis, and/or slower rates of involution (25, 26).

This is consistent with our findings that most of the associations between breast cancer susceptibility variants and risk-predicting mammographic density phenotypes are driven by associations with absolute or percent dense area. However, the recent discovery that at least one breast cancer susceptibility variant (at *NTN4*) appears to be solely associated with adjusted non-dense tissue provides evidence for the hypothesis that non-dense fatty tissue may play an independent role in the pathogenesis of breast cancer (27);

but we did not find mediation of the *NTN4*-breast cancer association by non-dense area in our study. Further, the fact that the association of adjusted dense area with risk is in the opposite direction to that for adjusted non-dense area, and these two measures are negatively correlated, raises the possibility that the risk associations with adjusted non-dense area and adjusted dense area are just the “opposite sides of the same coin” (6). In this regard, it is interesting that the familial correlations are very similar for non-dense area, as they are for adjusted percent dense area (28). Studies that have examined whether adjusted absolute non-dense area is independently associated with breast cancer risk have produced contradictory results (6, 29-32).

Variants in the regions of *MKL1*, *ESR1* and *ZNF365* were among nine variants previously reported to be associated with bra cup size, although none of the variants overlap those examined in this study (33). The *ESR1* variant reported by Eriksson *et al.* is in moderate LD with both the *ESR1* variants reported on here ($r^2=0.23$ with rs2046210 and with rs375318, although the mutual LD between these is just 0.07), and Eriksson’s *MKL1* variant (rs73167017) is also in moderate LD with *MKL1* rs6001930 reported here ($r^2=0.33$), but the *ZNF365* variant, rs7089814, is independent of rs10995190 ($r^2=0.04$). Within *MKL1*, rs6001930 is correlated with rs5995871 ($r^2=0.75$), which has been recently reported to be associated with mammographically-measured female breast size, which correlates strongly with non-dense area (34). Little is known about the functionality of any of the density-associated variants identified in this study.

Our study benefited from its large sample size and genotyping performed using the same custom Illumina iSelect genotyping array in 11 of the 13 studies. We also used the same strict quality controls for all studies except the two GWAS, and exclusion of data from these two GWAS-based studies did not substantially affect the findings reported here (data

not shown). Mammographic density measurements were performed using well-established methods shown to have high repeatability by trained observers, with all analyses being adjusted for study to reduce the impact of any between-study differences in the type of films available, digitizer used, quality of the density readings, and source of covariate data and other unmeasured confounders. We reproduced established associations between the three risk-predicting density phenotypes and measured breast cancer risk factors as well as prior genetic associations with breast cancer variants. Since over 80% of our sample population were postmenopausal, these results are generally applicable to postmenopausal women. However, based on previous work using longitudinal twin data, we have shown that the familial/genetic component of mammographic density is established before mid-life (35) and therefore, we believe that the direction of the associations reported in this study would be the same for premenopausal women.

More than 40 studies have found an association between mammographic density and breast cancer risk, many using different qualitative or quantitative methods of measuring mammographic density (19, 36). This suggests that mammographic density, as currently measured, is a useful biomarker. Our previous collaborations (9, 37) have demonstrated that data from multiple mammographic density studies can be combined to produce internally consistent results. One reason for this is the very wide variation in mammographic density measures within populations, even for women of the same age and BMI.

In summary, our findings provide further support for shared genetic determinants of breast cancer risk and the mammographic density measures that predict risk, presumably representing shared etiological pathways. While the contributions of the genetic risk markers identified to date explain little of the phenotypic variance, uncovering the cause of

familial aggregation (the so-called “missing heritability”) of the mammographic density measures that predict breast cancer could substantially increase understanding of the biological pathways involved in the development of the disease.

Acknowledgements

This study would not have been possible without the contributions of the following: Per Hall (COGS); Douglas F. Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Douglas F. Easton, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility.

References

1. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2006;15:1159-69.
2. Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, McCredie MR, et al. Heritability of mammographic density, a risk factor for breast cancer. *N Engl J Med.* 2002;347:886-94.
3. Stone J, Dite GS, Gunasekara A, English DR, McCredie MR, Giles GG, et al. The heritability of mammographically dense and nondense breast tissue. *Cancer Epidemiol Biomarkers Prev.* 2006;15:612-7.
4. Olson JE, Sellers TA, Scott CG, Schueler BA, Brandt KR, Serie DJ, et al. The influence of mammogram acquisition on the mammographic density and breast cancer association in the Mayo Mammography Health Study Cohort. *Breast Cancer Res.* 2012;14:R147.
5. Martin LJ, Melnichouk O, Guo H, Chiarelli AM, Hislop TG, Yaffe MJ, et al. Family history, mammographic density, and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2010;19:456-63.
6. Baglietto L, Krishnan K, J S, Apicella C, English DR, Hopper J, et al. Associations of mammographic dens and non-dense area and body mass index with risk of breast cancer. *Am J Epidemiol.* 2014;179:475-83.
7. Odefrey F, Stone J, Gurrin LC, Byrnes GB, Apicella C, Dite GS, et al. Common genetic variants associated with breast cancer and mammographic density measures that predict disease. *Cancer Res.* 2010;70:1449-58.

8. Vachon CM, Sellers TA, Carlson EE, Cunningham JM, Hilker CA, Smalley RL, et al. Strong evidence of a genetic determinant for mammographic density, a major risk factor for breast cancer. *Cancer Res.* 2007;67:8412-8.
9. Vachon CM, Scott CG, Fasching PA, Hall P, Tamimi RM, Li J, et al. Common breast cancer susceptibility variants in LSP1 and RAD51L1 are associated with mammographic density measures that predict breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2012;21:1156-66.
10. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet.* 2010;42:504-7.
11. 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.
12. Varghese JS, Thompson DJ, Michailidou K, Lindstrom S, Turnbull C, Brown J, et al. Mammographic breast density and breast cancer: evidence of a shared genetic basis. *Cancer Res.* 2012;72:1478-84.
13. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet.* 2013;45:353-61.
14. 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.
15. 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 epidemiology, biomarkers & prevention* : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008;17:3509-16.
16. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. *Phys Med Biol*. 1994;39:1629-38.
 17. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer. *Breast Cancer Res*. 2005;7:R854-61.
 18. Zhou C, Chan HP, Petrick N, Helvie MA, Goodsitt MM, Sahiner B, et al. Computerized image analysis: estimation of breast density on mammograms. *Med Phys*. 2001;28:1056-69.
 19. McCormack VA, Highnam R, Perry N, dos Santos Silva I. Comparison of a new and existing method of mammographic density measurement: intramethod reliability and associations with known risk factors. *Cancer Epidemiol Biomarkers Prev*. 2007;16:1148-54.
 20. Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med*. 1997;16:1515-27.
 21. Fisher R. *Statistical Methods for Research Workers*. 14th ed. New York: Hafner/MacMillan; 1970.
 22. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met*. 1996;58:267-88.
 23. Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, et al. Digital mammographic density and breast cancer risk: a case inverted question mark control

- study of six alternative density assessment methods. *Breast Cancer Res.* 2014;16:439.
24. Pinto Pereira SM, McCormack VA, Hipwell JH, Record C, Wilkinson LS, Moss SM, et al. Localized fibroglandular tissue as a predictor of future tumor location within the breast. *Cancer Epidemiol Biomarkers Prev.* 2011;20:1718-25.
 25. Ghosh K, Hartmann LC, Reynolds C, Visscher DW, Brandt KR, Vierkant RA, et al. Association between mammographic density and age-related lobular involution of the breast. *J Clin Oncol.* 2010;28:2207-12.
 26. Ginsburg OM, Martin LJ, Boyd NF. Mammographic density, lobular involution, and risk of breast cancer. *Br J Cancer.* 2008;99:1369-74.
 27. DeFilippis RA, Chang H, Dumont N, Rabban JT, Chen YY, Fontenay GV, et al. CD36 repression activates a multicellular stromal program shared by high mammographic density and tumor tissues. *Cancer discovery.* 2012;2:826-39.
 28. Nguyen TL, Schmidt DF, Makalic E, Dite GS, Stone J, Apicella C, et al. Explaining variance in the cumulus mammographic measures that predict breast cancer risk: a twins and sisters study. *Cancer Epidemiol Biomarkers Prev.* 2013;22:2395-403.
 29. Lokate M, Peeters PH, Peelen LM, Haars G, Veldhuis WB, van Gils CH. Mammographic density and breast cancer risk: the role of the fat surrounding the fibroglandular tissue. *Breast Cancer Res.* 2011;13:R103.
 30. 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:R100.

31. Stone J, Ding J, Warren RM, Duffy SW, Hopper JL. Using mammographic density to predict breast cancer risk: dense area or percentage dense area. *Breast Cancer Res.* 2010;12:R97.
32. Pettersson 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.
33. 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 medical genetics.* 2012;13:53.
34. 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;50:666-73.
35. Stone J, Dite GS, Giles GG, Cawson J, English DR, Hopper JL. Inference about causation from examination of familial confounding: application to longitudinal twin data on mammographic density measures that predict breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2012;21:1149-55.
36. Vachon CM, Ghosh K, Brandt KR. Mammographic density: potential as a risk factor and surrogate marker in the clinical setting. *Curr Breast Cancer Rep.* 2013;5:183-93.
37. 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;43:185-7.
38. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational,

- interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res.* 2004;6:R375-89.
39. Hopper JL, Chenevix-Trench G, Jolley DJ, Dite GS, Jenkins MA, Venter DJ, et al. Design and analysis issues in a population-based, case-control-family study of the genetic epidemiology of breast cancer and the Co-operative Family Registry for Breast Cancer Studies (CFRBCS). *J Natl Cancer Inst Monogr.* 1999:95-100.
40. Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MR, et al. Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. *J Natl Cancer Inst.* 2003;95:448-57.
41. Heusinger K, Loehberg CR, Haeberle L, Jud SM, Klingsiek P, Hein A, et al. Mammographic density as a risk factor for breast cancer in a German case-control study. *Eur J Cancer Prev.* 2011;20:1-8.
42. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, et al. EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer. Br J Cancer.* 1999;80 Suppl 1:95-103.
43. Kelemen LE, Wang X, Fredericksen ZS, Pankratz VS, Pharoah PD, Ahmed S, et al. Genetic variation in the chromosome 17q23 amplicon and breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2009;18:1864-8.
44. Giles GG, English DR. The Melbourne Collaborative Cohort Study. *IARC Sci Publ.* 2002;156:69-70.
45. Woolcott CG, Maskarinec G, Haiman CA, Verheus M, Pagano IS, Le Marchand L, et al. Association between breast cancer susceptibility loci and mammographic density: the Multiethnic Cohort. *Breast Cancer Res.* 2009;11:R10.

46. Maskarinec G, Pagano I, Lurie G, Wilkens LR, Kolonel LN. Mammographic density and breast cancer risk: the multiethnic cohort study. *Am J Epidemiol.* 2005;162:743-52.
47. Tamimi RM, Colditz GA, Hankinson SE. Circulating carotenoids, mammographic density, and subsequent risk of breast cancer. *Cancer Res.* 2009;69:9323-9.
48. Tamimi RM, Cox DG, Kraft P, Pollak MN, Haiman CA, Cheng I, et al. 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.
49. Wedren S, Lovmar L, Humphreys K, Magnusson C, Melhus H, Syvanen AC, et al. Oestrogen receptor alpha gene haplotype and postmenopausal breast cancer risk: a case control study. *Breast Cancer Res.* 2004;6:R437-49.
50. Kataoka M, Antoniou A, Warren R, Leyland J, Brown J, Audley T, et al. Genetic models for the familial aggregation of mammographic breast density. *Cancer Epidemiol Biomarkers Prev.* 2009;18:1277-84.

Table 1. Design, sample size, data collection, mammographic characteristics and genotyping information for the 13 studies

Study name (reference)	Study Abbreviation	Design ^a	Number Cases/Controls	Source of covariate data			Film view ^b	Breast side ^c		Genotyping (GWAS or iCOGS)
				Reproductive variables	Anthropometry	Time between mammogram and data collection		Cases	Controls	
Australian Breast Cancer Family Study (38-40)	ABCFS	CC Family	103/0	Questionnaire	Self report	Within 3 years	CC	Contra	n/a	iCOGS
Bavarian Breast Cancer Cases and Controls (41)	BBCC	CC	512/367	Questionnaire	Self report	Within 30 days	CC	Contra	Average	iCOGS
European Prospective Investigation into Cancer (42)	EPIC	Cohort	86/968	Questionnaire	Measured	3 years prior	MLO	Contra	Average	iCOGS
Mayo Clinic Breast Cancer Study (43)	MCBCS	CC	677/864	Questionnaire	Measured	Within 30 days	CC	Contra	L	iCOGS
Melbourne Collaborative Cohort Study (44)	MCCS	Nested CC	68/28	Questionnaire	Measured	3 years prior	CC	R	R	iCOGS
Multiethnic Cohort Study (45, 46)	MEC	Nested CC	110/101	Questionnaire	Self report	Within 5 years prior	CC	Average ^d	Average	iCOGS
Old Amish Study		Family	0/400	Questionnaire	Measured	Within 30 days	CC	n/a	L or R	GWAS
Mayo Mammography Health Study (4)	MMHS	Nested CC	456/1166	Questionnaire	Measured	Within 30 days	CC	Average ^d	Average	iCOGS
Norwegian Breast Cancer Study	NBCS	CS	0/38	Questionnaire	Self report	Within 14 days	CC	n/a	L	iCOGS
Nurses Health Study (47, 48)	NHS	Nested CC	850/849	Questionnaire	Self report	Within 2 years	CC	Average ^d	Average	GWAS

Breast Cancer Susceptibility Variants and Mammographic Density

Ontario Familial Breast Cancer Registry (38)	OFBCR	Family	73/0	Questionnaire	Self report	2-9 years prior	CC	Contra	n/a	iCOGS
Singapore and Sweden Breast Cancer Study (49)	SASBAC	CC	869/783	Questionnaire	Self report	Mean 1 year post	MLO	Contra	L or R	iCOGS
Sisters in Breast Cancer Screening (50)	SIBS	Family	0/1359	Questionnaire	Measured	Within 1 year prior	MLO	n/a	Average	iCOGS

^aCC=case-control study; CS=cross-sectional study

^bCC=cranio-caudal view; MLO= medio-lateral oblique view

^cAverage=average from left and right breasts; contra=unaffected contra-lateral breast; L=left breast; n/a=not applicable; R= right breast

^dPrediagnostic films

Breast Cancer Susceptibility Variants and Mammographic Density

Table 2. Summary characteristics at time of mammogram and by case status for the participating studies

Characteristic	Category	Breast cancer cases		Non-cases	
		N	%	N	%
Age (yrs)	<50	542	14.3	776	11.2
	50-59	1197	31.5	2307	33.3
	≥ 60	2065	54.3	3840	55.5
Parity	Nulliparous	466	12.3	775	11.2
	Parous	3283	86.3	6063	87.6
	Unknown	55	1.5	85	1.2
Menopausal status	Pre-menopausal	611	16.1	1268	18.3
	Post-menopause	3152	82.9	5575	80.5
	Unknown	41	1.1	80	1.2
Postmenopausal hormone therapy use	Ever	1848	48.6	2942	42.5
	Never	1710	45.0	2885	41.7
	Unknown	246	6.5	1096	15.8
BMI (kg/m ²)	<25	1644	43.2	2751	39.7
	≥ 25	2121	55.8	4122	59.5
	Unknown	39	1.0	50	0.7
Mammographic side, view ^a	L – CC	426	11.2	1102	15.9
	R - CC	511	13.4	296	4.3

Breast Cancer Susceptibility Variants and Mammographic Density

LR average - CC	1911	50.2	2415	34.9
L - MLO	446	11.7	415	6.0
R - MLO	510	13.4	409	5.9
LR average - MLO	0	0	2286	33.0

^aCC=cranio-caudal; L=left; MLO= medio-lateral oblique; R=right

Breast Cancer Susceptibility Variants and Mammographic Density

Table 3. Top associations between common breast cancer susceptibility variants and each of the mammographic measures

SNP	Locus	Alleles ^a	Percent dense area			Dense area			Non-dense area		
			Beta Estimate ^b	Standard Error	P-value	Beta Estimate ^b	Standard Error	P-value	Beta Estimate ^b	Standard Error	P-value
<i>Previously reported:</i>											
rs10995190	ZNF365	A/G	0.16	0.03	8.5E-09	0.25	0.04	4.7E-11	-0.005	0.04	0.91
rs3817198	LSP1	A/G	0.09	0.02	4.4E-05	0.16	0.03	1.3E-07	-0.001	0.03	0.97
rs2046210	ESR1	G/A	0.10	0.02	2.4E-06	0.14	0.03	1.7E-06	-0.02	0.03	0.5
<i>Novel associations:</i>											
rs6001930	MKL1	A/G	-0.06	0.03	0.044	-0.18	0.04	3.2E-05	-0.23	0.05	1.7E-06
rs1432679	EBF1	A/G	0.09	0.02	1.1E-05	0.09	0.03	7.1E-04	-0.11	0.03	4.5E-04
rs17356907	NTN4	G/A	0.03	0.02	0.16	-0.01	0.03	0.68	-0.12	0.03	2.4E-04
rs17817449	MIR1972-2:FTO	C/A	0.07	0.02	4.4E-04	0.09	0.03	0.001	-0.06	0.03	0.06
rs3757318	ESR1	G/A	0.07	0.04	0.066	0.19	0.05	4.6E-04	0.12	0.06	5.0E-02
rs12710696	2p24.1	G/A	-0.07	0.02	8.7E-04	-0.10	0.03	5.9E-04	0.03	0.03	0.32

^aSecond allele is modeled allele (breast cancer risk allele).

^bOrdinal per risk allele estimate, Age, 1/BMI, study adjusted.

Bold type denotes SNPs with $p < 6.5 \times 10^{-4}$ for that association. Study heterogeneity $p > 0.05$ for all SNPs apart from the association between rs2046210 and adjusted dense area ($p = 0.03$).

Table and Figure Legends

Table 1. Design, sample size, data collection, mammographic characteristics and genotyping information for the 13 studies

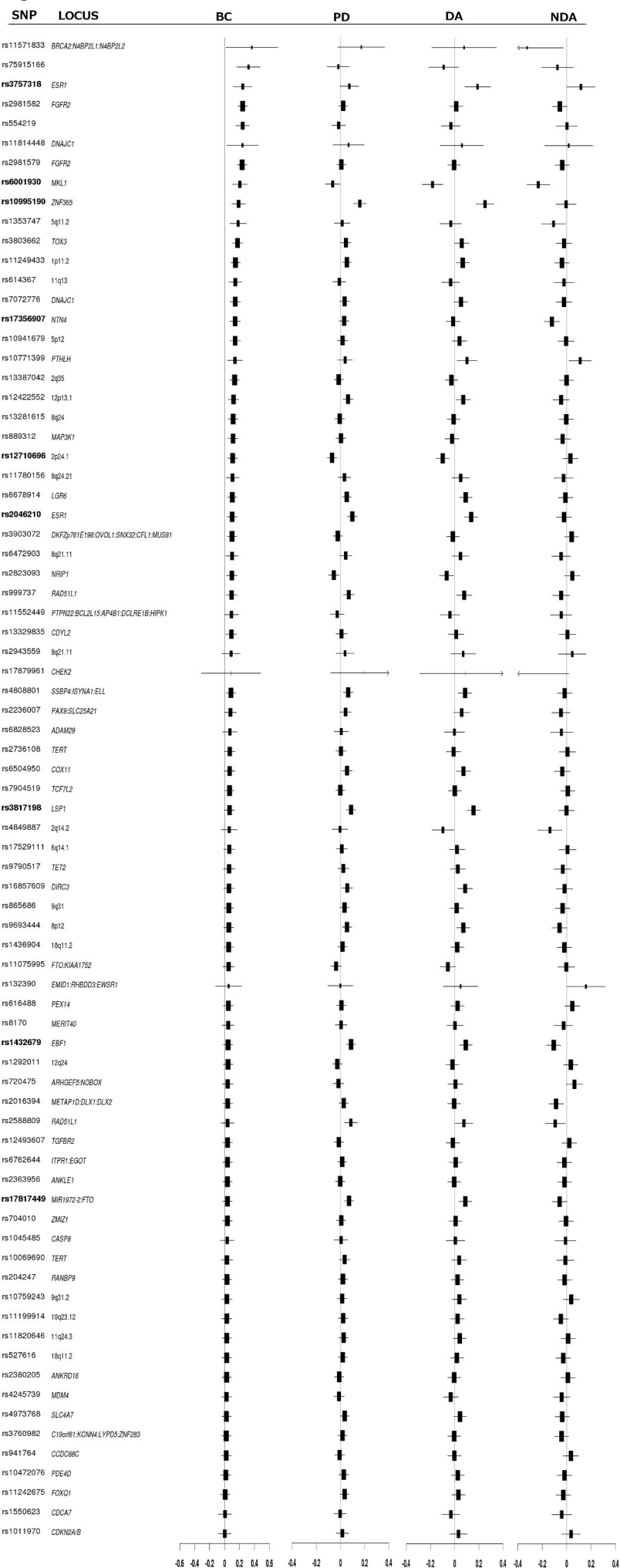
Table 2. Summary characteristics of the participating studies

Table 3. The top associations between common breast cancer susceptibility variants and each of the risk-predicting mammographic measures

Figure 1. Associations between the 77 common breast cancer susceptibility SNPs and breast cancer (BC), adjusted percent density (PD), adjusted dense area (DA) and adjusted non-dense area (NDA), ordered by the magnitude of the association with breast cancer

Figure 2. QQ plots before and after exclusion of the top 14 breast cancer susceptibility SNPs most strongly associated with the mammographic density measures: (a) percent density; (b) dense area; (c) non-dense area

Figure 1



Ordinal per risk allele estimates shown for each phenotype. All models adjusted for age, 1/BMI, and study effects.

Figure 2