Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation	Lindström, S., D. J. Thompson, A. D. Paterson, J. Li, G. L. Gierach, C. Scott, J. Stone, et al. 2015. "Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk." Nature communications 5 (1): 5303. doi:10.1038/ncomms6303. http://dx.doi.org/10.1038/ncomms6303.
Published Version	doi:10.1038/ncomms6303
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:15034790
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk

Sara Lindström[1,2], Deborah J. Thompson[3,4], Andrew D. Paterson[5], Jingmei Li[6], Gretchen L. Gierach[7], Christopher Scott[8], Jennifer Stone[9], Julie A. Douglas[10], Isabel dos-Santos-Silva[11], Pablo Fernandez-Navarro[12,13], Jajini Verghase[3,4,14], Paula Smith[3,4], Judith Brown[3,4], Robert Luben[3], Nicholas J. Wareham[15], Ruth J.F. Loos[15,16], John A. Heit[17], V. Shane Pankratz[8], Aaron Norman[8], Ellen L. Goode[8], Julie M. Cunningham[18], Mariza deAndrade[8], Robert A. Vierkant[8], Kamila Czene[19], Peter A. Fasching[20,21], Laura Baglietto[22,23], Melissa C. Southey[24], Graham G. Giles[22,23], Kaanan P. Shah[10], Heang-Ping Chan[25], Mark A. Helvie[25], Andrew H. Beck[26], Nicholas W. Knoblauch[26], Aditi Hazra[1,2,27], David J. Hunter[1,2,27], Peter Kraft[1,2,28], Marina Pollan[12,13], Jonine D. Figueroa[7], Fergus J. Couch[8,18], John L. Hopper[23], Per Hall[19], Douglas F. Easton[3,4,29], Norman F. Boyd[30], Celine M. Vachon[8,†], and Rulla M. Tamimi[1,2,27,†]

[1]Program in Genetic Epidemiology and Statistical Genetics, Harvard School Of Public Health, Boston, MA, 02115, USA

[2]Department of Epidemiology, Harvard School Of Public Health, Boston, MA, 02115, USA

[3]Centre for Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK

[4]Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK

[5]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada

[6]Human Genetics, Genome Institute of Singapore, Singapore, 138672, Singapore

[7]Hormonal and Reproductive Epidemiology Branch, National Cancer Institute, Bethesda, MD 20850, USA

[8]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA

Correspondence to: Rulla Tamimi, Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA, rulla.tamimi@channing.harvard.edu.
†These authors jointly supervised this work

[9]Centre for Genetic Origins of Health and Disease, University of Western Australia, Perth, WA 6009, Australia

[10]Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109, USA

[11]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

[12]Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid 28029, Spain

[13]Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública – CIBERESP) 28029, Spain

[14]Plastic Surgery Unit, Royal Free Hospital, London, UK

[15]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge CB1 8RN, UK

[16]The Icahn School of Medicine at Mount Sinai, The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, New York, NY 10029, USA

[17]Division of Cardiovascular Disease, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA

[18]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA

[19]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden

[20]Department of Gynecology and Obstetrics, Erlangen University Hospital, Friedrich Alexander University of Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, 910 54 Erlangen, Germany

[21]University of California at Los Angeles, Department of Medicine, Division Hematology/ Oncology, David Geffen School of Medicine, Los Angeles, CA 90024, USA

[22]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne 3004, Australia

[23]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, 3010, Australia

[24]Department of Pathology, University of Melbourne, Melbourne 3010, Australia

[25]Department of Radiology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

[26]Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02115, USA

[27]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

[28]Department of Biostatistics, Harvard School Of Public Health, Boston, MA 02115, USA

[29]Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK

[30]Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute, Toronto, Ontario M5G 2M9, Canada

Abstract

Mammographic density reflects the amount of stromal and epithelial tissues in relation to adipose tissue in the breast and is a strong risk factor for breast cancer. Here we report the results from meta-analysis of genome-wide association studies (GWAS) of three mammographic density phenotypes: dense area, non-dense area and percent density in up to 7,916 women in stage 1 and an additional 10,379 women in stage 2. We identify genome-wide significant ($P<5\times10^{-8}$) loci for dense area (AREG, ESR1, ZNF365, LSP1/TNNT3, IGF1, TMEM184B, SGSM3/MKL1), non-dense area (8p11.23) and percent density (PRDM6, 8p11.23, TMEM184B). Four of these regions are known breast cancer susceptibility loci, and four additional regions were found to be associated with breast cancer ($P<0.05$) in a large meta-analysis. These results provide further evidence of a shared genetic basis between mammographic density and breast cancer and illustrate the power of studying intermediate quantitative phenotypes to identify putative disease susceptibility loci.

INTRODUCTION

Variations in the appearance of the mammogram reflect differences in breast fibroglandular tissue that appears white or radio-dense, and fat that appears black or non-dense. After adjustment for age and body mass index (BMI), the proportion of the total breast area that is dense (percent density (PD)) is a strong risk factor for breast cancer[1], and both dense (DA) and non-dense areas (NDA), are also independently associated with breast cancer risk[2,3]. PD, DA and NDA are all highly heritable (0.6–0.7)[4,5], but to date few genetic loci associated with mammographic density have been identified[6–8].

Here we report results from a two-stage (discovery and replication stages) GWAS of DA, NDA and PD, respectively. We identify genome-wide significant ($P<5\times10^{-8}$) loci for dense area (*AREG*, *ESR1*, *ZNF365*, *LSP1/TNNT3*, *IGF1*, *TMEM184B* and *SGSM3/MKL1*), non-dense area (8p11.23) and percent density (*PRDM6*, 8p11.23, *TMEM184B*). Our results add to the growing body of evidence that mammographic density and breast cancer risk share a genetic component.

RESULTS

Our discovery phase included eleven studies with GWAS data (Methods, Supplementary Note 1) comprising a total of 7,916 women. Study subjects were predominantly post-menopausal women of European ancestry participating in the Markers of Density (MODE) consortium. Mammographic density was measured using CUMULUS[9] (Supplementary Table 1) and 1,642 (21%) of the subjects were breast cancer cases. All studies were imputed to HapMap phase II before meta-analysis (Supplementary Table 2). For each SNP, we combined study-specific p-values and direction of association using the METAL software[10]. We assessed 200 promising SNPs for replication in up to 10,379 women from eleven different studies (Supplementary Table 3, Supplementary Note 2).

For DA (n=7,600), no SNP reached genome-wide significance in the discovery phase (Supplementary Figs. 1 and 2). However, through replication analysis (Supplementary Table 4), we identified seven independent loci significantly associated ($P<5\times10^{-8}$) with DA (Table 1, Supplementary Figs. 3 and 4) including *AREG, ESR1, ZNF365, LSP1/TNNT3, IGF1, TMEM184B* and *SGSM3/MKL1*.

The *AREG* gene is a member of the epidermal growth factor family that promotes growth of normal epithelial cells and variants strongly correlated with our top SNP rs10034692 in this region have previously been associated with breast size[11]. Although we observed the strongest association for rs10034692, another SNP (rs12642133) located 116kb away and in weak linkage disequilibrium (LD) with rs10034692 (r-sq=0.16, D'=1.00) also reached genome-wide significance (Supplementary Table 4). We investigated these two SNPs further in 6,624 women from the NHS, BBCC, MCBCS and MMHS studies for whom we had individual-level genotype data. Both SNPs were associated with DA in this dataset when analyzed separately ($\beta=-0.16$, $P=0.0002$ for rs10034692 and $\beta=0.17$, $P=9\times10^{-6}$ for rs12642133). Including both SNPs in the same model attenuated the signal for both SNPs ($\beta=-0.10$, $P=0.04$ for rs10034692 and $\beta=0.13$, $P=0.002$ for rs12642133). Thus, it is possible that these two SNPs are either a proxy for another yet unidentified causal SNP or that they represent two independent causal SNPs. Interestingly, rs12642133 is located in a weak enhancer region in human mammary epithelial cells (HMEC).

SNPs in *ESR1* have earlier been associated with breast cancer risk[12–15] and rs12665607 identified here is in strong LD with the breast cancer SNP rs3757318 (r-sq=0.87, D'=1.00) and in moderate LD with SNPs previously associated with breast size[11].

The rs10995190 SNP in the *ZNF365* region has been associated with both PD[6] and breast cancer risk[14] but this is the first time it has been found to be associated with DA specifically. We observed multiple SNPs in the *ZNF365* gene associated with DA and since multiple independent SNPs in *ZNF365* are associated with breast cancer[14,16], we conducted conditional analyses to identify potential independent signals. In particular, SNPs rs1949359 (r-sq=0.08, D'=0.36 with rs10995190) and rs10733779 (r-sq=0.11, D'=1.00 with rs10995190) showed genome-wide significant associations with DA. After adjusting for rs10995190, the associations for both rs1949359 ($P=4.4\times10^{-5}$ before and $P=0.008$ after adjustment) and rs10733779 ($P=1.9\times10^{-6}$ before and $P=0.002$ after adjustment) were attenuated. Additional analyses in larger datasets will be necessary to determine if there are multiple independent SNPs in this region.

We identified a rare (minor allele frequency (MAF)=0.02) SNP 222 kb upstream of *IGF1* that was associated with DA. *IGF1* is a candidate gene for breast cancer risk[17] and is hypothesized to be involved in breast development. Indeed, circulating levels of IGF-1 are associated with breast cancer risk[18].

We also confirmed previous findings[8] that rs3817198 in the known breast cancer gene *LSP1* is associated with DA and also observed a genome-wide significant association for a weakly correlated SNP rs909116 (r-sq=0.24, D'=0.82). Both these SNPs have been associated with breast cancer risk and the recently published iCOGS[19] analysis of breast cancer found that

rs3817198 is the SNP most strongly associated with breast cancer at the *LSP1* locus. Large-scale fine-mapping efforts are needed to pinpoint the causal variant(s).

SNP rs7289126 (*TMEM184B)* was associated with both DA and PD. A correlated SNP rs738322 (r-sq=0.34, D'=0.71) located in the *PLA2G6* gene has previously been associated with cutaneous nevi[20]. Interestingly, two recent independent studies recently reported a link between cutaneous nevi and breast cancer[21,22] and it is possible that this link can be partly explained through a shared genetic origin between cutaneous nevi and mammographic density.

The SNP rs17001868 (*SGSM3/MKL1* region) is in moderate LD (r-sq=0.41, D'=0.76) with rs6001930 that has been previously associated with breast cancer[19]. We also observed several nearby SNPs located in the *TNRC6B* and *MKL1* genes that were associated with DA. However, these SNPs did not remain significant after adjusting for rs17001868.

For NDA (n=7,600), multiple SNPs at 8p11.23 reached genome-wide significance in the discovery phase (Supplementary Figs. 5 and 6); this region has previously been associated with breast size[11,23] (Table 1, Supplementary Figs. 8 and 9). Replication analysis (Supplementary Table 5) confirmed this region (top SNP rs7816345, combined $P=2.4\times10^{-23}$), and this SNP was also associated with PD on a genome-wide significant level.

For PD (n=7,916), the only two regions that reached genome-wide significance in the discovery stage were the previously identified *ZNF365*[6] and 12q24[7] loci (Supplementary Figs. 10 and 11). Through replication analysis (Supplementary Table 6), we identified three new loci ($P<5\times10^{-8}$) that mapped to *PRDM6*, 8p11.23 and *TMEM184B* (Table 1, Supplementary Figs 12 and 13). rs7816345 (8p11.23) was also significantly associated with NDA and rs7289126 (*TMEM184B*) with DA on a genome-wide significance level. SNP rs186749 is located in *PRDM6*, a gene involved in regulation of endothelial cell proliferation, survival and differentiation. Interestingly, we observed a borderline association ($P=2.6\times10^{-7}$) between rs186749 and DA (Supplementary Table 4). We also observed two SNPs in *ZNF365*, rs10733779 and rs10509168, that reached genome-wide significance but their associations were attenuated when adjusting for the known PD SNP rs10995910. As with DA, analysis in larger datasets will be needed to assess the possibility of multiple independent SNPs in this region.

We used data from the ENCODE[24] project to identify potential overlap between SNPs in regions associated with mammographic density phenotypes and regulatory elements in mammary tissue (Supplementary Table 7). We identified multiple SNPs in these regions that were in strong LD (r-sq 0.8) with the lead SNPs and mapped to regulatory regions as defined by DNAse I hypersensitive site (DHS) or enhancer histone marks in mammary tissue for the *ESR1*, *IGF1*, *TMEM184B*, *SGSM3/MKL1* and 8p11.23 regions. In particular, several SNPs including rs77275268 (proxy for rs12665607) in the *ESR1* region map to a DHS in the breast MCF-7 and HMEC cell lines. SNP rs77275268 has previously been shown to disrupt a partially methylated CpG sequence within a known CTCF binding site[25]. Interestingly both rs77275268 and rs4820328 (proxy for rs7289126) in the *TMEM184B*

region are in regions that bind CTCF. CTCF is believed to play genome-wide role in transcriptional regulation and chromatic structure. In addition, rs4820328 also mapped to enhancer histone marks and DHS in HMEC cell lines. Based on these data, rs4820328 and rs77275268 are intriguing candidates for further follow up. We also identified SNPs in these regions that bind several proteins implicated in breast cancer including GATA3, ESR1, FOXA1, YY1, RAD21, SMC3, GR and EGR1. To explore potential function of identified SNPs further, we assessed their association with gene expression levels in adipose tissue and lymphoblastoid cell lines (LCL)[26]. The DA and PD SNP rs7289126 (*TMEM184B*) was associated with expression of MAFF and ANKRD54 in LCL (P<0.001) and BAIAP2L2 in adipose tissue (P<0.00001). rs17001868 (*SGSM3/MKL1*) was associated with SGSM3 expression in both adipose tissue and LCL (P<0.0001). We also examined if any of these SNPs (or proxies) were associated with transcript levels in breast cancer tumors using data from The Cancer Genome Atlas[27] (TCGA). We conducted both cis (within 1 Mb of the transcription start or end site) and trans (genome-wide) eQTL analysis. Although we did not identify any significant pathways in gene-set enrichment analysis, we identified some significant eQTLs with a raw p<0.00024 (Supplementary Table 9). Interestingly, rs4820328 in the *TMEM184B* region that showed up in the ENCODE analysis was also associated with multiple transcript levels in TCGA.

To investigate if SNPs associated with mammographic density phenotypes are also associated with breast cancer, we accessed data from the GAME-ON (http:// gameon.dfci.harvard.edu) and iCOGS breast cancer meta-analysis based on 62,533 cases and 60,976 controls (Table 2). Eight out of nine SNPs were associated with breast cancer risk (*P*<0.05), four of which have already been reported to be associated with breast cancer on a genome-wide significance level (*ESR1, ZNF365, LSP1* and *SGSM3/MKL1*)[12–15,19,28]. Four additional SNPs (*PRDM6*, 8p11.23, *IGF1* and *TMEM184B*) were nominally associated with breast cancer (P<0.05, Table 2) and indicate potential new breast cancer susceptibility loci. Among the eight SNPs associated with both mammographic density phenotypes and breast cancer, six SNPs showed consistent direction between the mammographic density and breast cancer association, whereas *SGSM3/MKL1* and 8p11.23 showed conflicting direction of associations with breast cancer in relation to the mammographic density association. We conducted SNP-breast cancer association analyses with and without adjusting for mammographic density (Supplementary Table 9) in up to 3,696 breast cancer cases and 4,768 controls for whom we had mammographic density data on. We did not observe strong evidence that mammographic density mediates the SNP-breast cancer association, but we note that our low sample size limits our ability to draw conclusions from these analyses.

The SNPs identified here explain only a small fraction of the variance of DA (1.0%), NDA (0.4%) and PD (0.6%). We generated phenotype-specific genotype scores and estimated the difference in density associated with each density-increasing allele carried. The score-specific differences per allele were 1.94 cm^2 for DA, 8.58 cm^2 for NDA and 0.77% for PD, respectively. It is noteworthy that two out of three SNPs associated with PD were associated with either DA or NDA and that there is overlap between our findings here and two recent GWAS of a correlated but distinct phenotype, breast size[11,23]. This was also partly reflected in our GWAS analyses (Supplementary Table 10, Supplementary Figs. 10–12).

DISCUSSION

In this two-stage GWAS of mammographic density phenotypes we identified genome-wide significant loci for all three phenotypes investigated: dense area, non-dense area and percent density. Four of the identified regions are known breast cancer susceptibility loci, and four additional regions were found to be associated with breast cancer (P<0.05) in a large meta-analysis. These four mammographic density SNPs represent putative novel breast cancer loci.

While the majority of density related SNPs we identified showed associations that were consistent in direction with breast cancer risk, there were two SNPs that were inconsistent. rs17001868 (*SGSM3/MKL1*) was strongly associated with both dense area and breast cancer risk but in opposite directions. Interestingly, there is accumulating data that MKL1 may have both tumor inhibiting and tumor promoting roles depending on the cellular context. Recently, it was shown that the MKL1 signaling pathway was activated in ER-cell lines and silenced in ER+ cell lines[29]. Additionally, the *MKL1* breast cancer SNP has been shown to be associated with triple negative breast cancer[30]. rs17001868 has also been associated with SGSM3 (involved in signal transduction pathway) expression in both LCL and adipose tissue. While these data suggest that this SNP influences expression levels that may affect breast cancer risk, it is unclear how well these tissues represent expression in normal breast tissue. The differing effects of this pathway dependent on the ER status of the tissue[29] suggest that understanding the cellular environment is important. Although the underlying biology is still not well understood it suggests that it is possible that rs17001868 affects mammographic dense area and breast cancer risk through different mechanisms associated with different target genes. The majority of women included in our study were postmenopausal at the time of the mammogram. This single assessment of breast density will reflect both the formation of dense tissue early in life, as well as, influences such as age-related and lactation related involution. The apparent opposing directions of this locus on dense area and breast cancer risk may suggest important biologic differences of the effect of this SNP on breast tissue and breast cancer risk by factors we are unable to assess in the current study (e.g, age, menopausal status). Similarly, rs7816345 was also associated with apparent opposing directions on non-dense area and breast cancer risk. Again, this may reflect true biologic differences over the life course. For example, it has been demonstrated that adiposity during early life is inversely associated with breast cancer[31], while postmenopausal BMI is positively associated with breast cancer[32].

There are some weaknesses with our study that should be mentioned. First, we used the HapMap project as imputation panel which prohibited us from assessing the contribution of rare variants. Future genetic studies of mammographic density phenotypes should use more dense imputation panels such as the 1000 Genomes[33] that will provide a more complete coverage of the genome. Moreover, it is possible that the causal variant(s) within each mammographic density GWAS region was not captured here. Pin-pointing the causal variants will require not only denser genotyping and/or sequencing of these regions but also larger sample sizes. Another weakness with our study is that it was not designed or adequately powered to test if mammographic density mediates SNP effects on breast cancer.

Future large studies with both mammographic density and breast cancer data should assess such mediation effects.

In summary, we report multiple loci associated with mammographic density phenotypes. We identified six DA-specific loci of which five showed an association with breast cancer and one PD-specific locus also associated with breast cancer. We also report an additional locus associated with DA, PD and breast cancer risk as well as a locus associated with NDA, PD and breast cancer risk. These results confirm previous observations that mammographic density phenotypes and breast cancer risk share genetic origin and biological pathways[34]. Despite the smaller sample size in this mammographic density GWAS (N=7,916 in the discovery and N=10,379 in the replication phase) compared with recent large-scale breast cancer studies (N=22,627 in the discovery and N=87,170 in the replication phase)[19], our ability to identify known as well as putative novel breast cancer loci by studying mammographic density phenotypes demonstrates the power of using quantitative intermediate phenotypes to discover new disease loci.

METHODS

Ethics Statement

Each study obtained informed consent from patients and had relevant ethics and institutional approvals from the following institutions. Brigham and Women's Hospital (NHS), Harvard School of Public Health (NHSII), Norwich District Ethics Committee (EPIC-Norfolk), Karolinska Institutet (SASBAC), Mayo Clinic (MBCFS, MAYO VTE, MCOCS, MMHS, MCBCS), University Health Network, Toronto, Canada (TOR), Eastern Multicentre Research Ethics Committee (SIBS), Instituto de Salud Carlos III (DDM-Spain), University of Melbourne (AMDTSS), University of Michigan and University of Maryland (OOA), The Cancer Council Victoria Ethic Commitee (MCCS), Friedrich-Alexander University Erlangen-Nuremberg (BBCC), NCI Special Studies Institutional Review Board (PBCS) and National Research Ethics Committee (NREC) East of England – Cambridge South (SEARCH).

Study design

We conducted a meta-analysis of twelve GWAS of mammographic density (Supplementary Note 1). For DA and NDA, we had GWAS data from eleven studies and for PD we had GWAS data from twelve studies. To follow up promising SNPs (p<0.0001) (Supplementary Tables 4–6), we conducted replication efforts using data from three different sources: iCOGS, iSelect and *in silico* look-ups in GWAS data. We assessed a total of 200 SNPs that showed suggestive associations with DA, NDA or PD for replication. We pursued replication of 114 SNPs that were included on the iCOGS[19] array and genotyped additional 86 SNPs in 3,832 women using a customized iSelect array. For the replication analysis, we also included data from the Old Order Amish (OOA, n=1,472) GWAS and for the DA analysis, the Australian MD Twins and Sisters Study (AMDTSS) GWAS (n=343).

Genotyping, quality control and imputation

Study participants were genotyped on various genotyping platforms, and standard quality control filters for call rate, Hardy-Weinberg equilibrium p-value, and other measures were applied to exclude individuals and genotyped SNPs. To generate a common set of SNPs for meta-analysis, all studies were imputed to HapMap phase II (Supplementary Table 2). Imputed genotype dosage values (estimated reference allele count with a fractional value ranging from 0 to 2.0) were generated for approximately 2.5 million SNPs. SNPs with an imputation quality score <0.8 (as defined by the RSQR_HAT value in MACH, the PROPER_INFO in IMPUTE and the information content (INFO) measure in PLINK)) or a minor allele frequency <0.01 were excluded.

GWAS analysis

Primary association analysis was performed separately within each study. All studies except the Toronto/Melbourne (TOR) and AMDTSS used linear regression assuming an additive inheritance model. For imputed SNPs, the estimated number of effect alleles (ranging from 0 to 2) was used as a covariate. To account for the family structure in Minnesota Breast Cancer Family Study (MBCFS) and Sisters in Breast Screening (SIBS), we used the "multic" package as implemented in R. Multic uses a linear mixed effects model, whereby the genetic relatedness among individuals is incorporated into the covariance structure of the random effects[35,36]. The relationships between subjects within the SIBS study were adjusted for using the mmscore option within ProbABEL, based on the estimated genomic kinship matrix[37]. The fixed effect is used for the tests of association and covariate adjustment. The TOR and AMDTSS used logistic regression where women in the 10% top percentile of percent mammographic density (TOR) or dense area (AMDTSS) were defined as "cases" and women in the bottom 10% percentile were defined as "controls". As the included data from the Nurses' Health Study (NHS) were generated using two different genotyping platforms, they were analyzed as two separate studies. Similarly, data from the Singapore and Sweden Breast Cancer Study (SASBAC) were obtained through two separate genotyping efforts and therefore analyzed separately. All studies adjusted their analysis for age and BMI. Additional study-specific adjustment factors are described in Supplementary Table 2. Study-specific genomic inflation factors ranged between 0.99 and 1.07.

Meta-analysis

Meta-analysis was based on summary statistics from the participating studies. For each SNP, we combined study-specific p-values and direction of association using the METAL software[10]. Weights were proportional to study-specific genomic inflation factors and sample size. To account for the extreme sampling scheme in the TOR study, we up-weighted the study with a scale factor of 3.51. For a SNP to be considered in the meta-analysis, we required genotyping data from at least 3,000 women. We used Cochran's Q statistic to test for heterogeneity across studies.

Replication analysis

Candidate SNPs were followed up through replication genotyping and *in silico* look-ups. We obtained replication data from three separate sources: through the iCOGS genotyping[19]

effort, through a customized iSelect array and through look-ups in GWAS of mammographic density in the OOA and AMDTSS studies (Supplementary Note 2). In total, we successfully pursued and obtained replication data for 200 SNPs (Supplementary Tables 4–6) selected based on their suggestive association (p<0.0001) with at least one of the density phenotypes. We also included the breast cancer SNPs rs10771399, rs1292011, rs909116 and rs2823093 since they were associated with at least one mammographic density phenotype at p<0.05. We extracted genotype data on 114 SNPs for 7,303 women for whom we had both iCOGS and mammographic density data. For SNPs that were not included on the iCOGS array but had a proxy (r-sq 0.80) on the iCOGS array, we included the proxy instead. We also genotyped additional 86 SNPs in 3,878 women from the Nurses' Health Study II (NHSII), the Mayo Mammography Health Study (MMHS), the Mayo Clinic Breast Cancer Study (MCBCS) and the Melbourne Collaborative Cohort Study (MCCS) using a customized iSelect array. We excluded subjects with call rates <95% (N=44) and two subjects (out of 204 included duplicates) that showed multiple discordances leaving 3,832 subjects for analysis. Remaining duplicates had concordance >99%. In addition, we also included association results from the OOA (n=1,472) and AMDTSS (n=343 for the DA analysis) GWAS where available. To account for the extreme sampling scheme in AMDTSS, we up-weighted this study with a scale factor of 3.51. In total, our replication sample size for SNPs included on the iCOGS array was 9,118 women and the sample size for SNPs included on the iSelect 5,647 women.

Assessment of regulatory functions for identified SNPs

We used the ENCODE[24] data to assess if any of the identified mammographic density SNPs or their proxies (r-sq 0.8 in 1000 Genomes CEU population) are located in regulatory regions. Look-ups were made using the HaploReg[38] and RegulomeDB[39] software. We also investigated if identified mammographic density SNPs or their proxies were associated with gene expression in *cis* in adipose tissue and lymphoblastoid cell lines (LCL) in the MuTHER[40] data by accessing the GeneVar[26] database. To further explore the regulatory properties of the mammographic density SNPs, we conducted eQTL analyses on mammographic density SNPs and their proxies (r-sq 0.8) using data from The Cancer Genome Atlas (TCGA). We identified eQTLs using BeQTL (manuscript under review, http://beqtl.org) that robustly assesses the association between SNP genotypes and mRNA transcript levels using linear regression with bootstrap. We assessed a total of 22 SNPs and a total of 18,985 transcripts among 608 estrogen receptor positive cases and 19,105 transcripts among 177 estrogen receptor negative cases. To robustly define the correlation between SNP genotype and gene expression level, the 95% confidence interval and median of the t-statistic for the correlation coefficient were estimated via statistical bootstrap. For the bootstrap procedure, case resampling was performed N*log(N) times where N is the total number of cases. We computed p-values from the median t-statistic obtained in linear regression. Functional gene set analysis was performed using DAVID[41,42] (http://david.abcc.ncifcrf.gov/) for the set of transcripts achieving a raw p-value less than 0.00024 in the eQTL analysis.

Breast cancer association analysis

We looked up the association between mammographic density SNPs and breast cancer in the iCOGS[19] + GAME-ON breast cancer GWAS meta-analysis. The GAME-ON meta-analysis[13,19,43,44] can be found at (http://gameon.dfci.harvard.edu) and is based on eleven breast cancer GWAS. In total, the reported breast cancer associations for the replicated mammographic density SNPs were based on 62,533 breast cancer cases and 60,976 controls. We conducted logistic regression analysis with and without adjustment for mammographic density including up to 3,696 breast cancer cases and 4,768 controls from the NHS, NHSII, MCBCS, MMHS, BBCC, SASBAC and MCCS studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Vachon CM, et al. Mammographic density, breast cancer risk and risk prediction. Breast cancer research: BCR. 2007; 9:217. [PubMed: 18190724]

2. Pettersson A, et al. Nondense mammographic area and risk of breast cancer. Breast cancer research: BCR. 2011; 13:R100. [PubMed: 22017857]

3. Pettersson A, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. J Natl Cancer Inst. 2014; 106

4. Boyd NF, et al. Heritability of mammographic density, a risk factor for breast cancer. The New England journal of medicine. 2002; 347:886–94. [PubMed: 12239257]

5. Stone J, et al. The heritability of mammographically dense and nondense breast tissue. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2006; 15:612–7.

6. Lindstrom S, et al. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. Nature genetics. 2011; 43:185–7. [PubMed: 21278746]

7. Stevens KN, et al. Identification of a novel percent mammographic density locus at 12q24. Human molecular genetics. 2012; 21:3299–305. [PubMed: 22532574]

8. Vachon CM, et al. Common breast cancer susceptibility variants in LSP1 and RAD51L1 are associated with mammographic density measures that predict breast cancer risk. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2012; 21:1156–66.

9. Byng JW, et al. Symmetry of projection in the quantitative analysis of mammographic images. European journal of cancer prevention: the official journal of the European Cancer Prevention Organisation. 1996; 5:319–27.

10. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]

11. Eriksson N, et al. Genetic variants associated with breast size also influence breast cancer risk. BMC medical genetics. 2012; 13:53. [PubMed: 22747683]

12. Long J, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. PLoS genetics. 2012; 8:e1002532. [PubMed: 22383897]

13. Siddiq A, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Human molecular genetics. 2012; 21:5373–84. [PubMed: 22976474]

14. Turnbull C, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nature genetics. 2010; 42:504–7. [PubMed: 20453838]

15. Zheng W, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nature genetics. 2009; 41:324–8. [PubMed: 19219042]

16. Cai Q, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Human molecular genetics. 2011; 20:4991–9. [PubMed: 21908515]

17. Hunter DJ, et al. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nature reviews. Cancer. 2005; 5:977–85.

18. Key TJ, Appleby PN, Reeves GK, Roddam AW. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. The lancet oncology. 2010; 11:530–42. [PubMed: 20472501]

19. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nature genetics. 2013; 45:353–61. 361e1–2. [PubMed: 23535729]

20. Nan H, et al. Genome-wide association study identifies nidogen 1 (NID1) as a susceptibility locus to cutaneous nevi and melanoma risk. Human molecular genetics. 2011; 20:2673–9. [PubMed: 21478494]

21. Zhang M, et al. Association between Cutaneous Nevi and Breast Cancer in the Nurses' Health Study: A Prospective Cohort Study. PLoS Med. 2014; 11:e1001659. [PubMed: 24915186]

22. Kvaskoff M, et al. Association between Melanocytic Nevi and Risk of Breast Diseases: The French E3N Prospective Cohort. PLoS Med. 2014; 11:e1001660. [PubMed: 24915306]

23. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45:353–61. 361e1–2. [PubMed: 23535729]

24. Consortium EP, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

25. Stacey SN, et al. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. PLoS Genet. 2010; 6:e1001029. [PubMed: 20661439]

26. Yang TP, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics. 2010; 26:2474–6. [PubMed: 20702402]

27. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]

28. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447:1087–93. [PubMed: 17529967]

29. Kerdivel G, et al. Activation of the MKL1/actin signaling pathway induces hormonal escape in estrogen-responsive breast cancer cell lines. Mol Cell Endocrinol. 2014; 390:34–44. [PubMed: 24721635]

30. Purrington KS, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis. 2014; 35:1012–9. [PubMed: 24325915]

31. Baer HJ, Tworoger SS, Hankinson SE, Willett WC. Body fatness at young ages and risk of breast cancer throughout life. Am J Epidemiol. 2010; 171:1183–94. [PubMed: 20460303]

32. van den Brandt PA, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol. 2000; 152:514–27. [PubMed: 10997541]

33. Genomes Project C, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

34. Varghese JS, et al. Mammographic breast density and breast cancer: evidence of a shared genetic basis. Cancer research. 2012; 72:1478–84. [PubMed: 22266113]

35. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. American journal of human genetics. 1998; 62:1198–211. [PubMed: 9545414]

36. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. American journal of human genetics. 1994; 54:535–43. [PubMed: 8116623]

37. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC bioinformatics. 2010; 11:134. [PubMed: 20233392]

38. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012; 40:D930–4. [PubMed: 22064851]

39. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome research. 2012; 22:1790–7. [PubMed: 22955989]

40. Nica AC, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS genetics. 2011; 7:e1002003. [PubMed: 21304890]

41. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009; 4:44–57.

42. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009; 37:1–13. [PubMed: 19033363]

43. Ghoussaini M, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nature genetics. 2012; 44:312–8. [PubMed: 22267197]

44. Garcia-Closas M, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nature genetics. 2013; 45:392–8. 398e1–2. [PubMed: 23535733]

Table 1

SNPs associated with mammographic dense area (DA), non-dense area (NDA) and percent density (PD).

Lead SNP	Chr: Position[1]	Gene[2]	Alleles[3]	MAF[4]	Stage	Z-Score	P	Stage 1+2			
								β^5 (se)	Z-Score	P	P (Het)[6]
SNPs associated with mammographic dense area											
rs10034692	4:75638651	AREG	A/G	0.26	1	−4.67	3.00×10^{-6}	−0.16 (0.03)	−6.35	2.12×10^{-10}	0.77
					2	−4.44	9.06×10^{-6}				
rs12665607	6:151988322	ESR1	T/A	0.08	1	3.91	9.16×10^{-5}	0.17 (0.04)	5.64	1.71×10^{-8}	0.27
					2	4.07	4.62×10^{-5}				
rs10995190	10:63948688	ZNF365	G/A	0.16	1	−4.65	3.36×10^{-6}	−0.24 (0.03)	−8.26	1.49×10^{-16}	0.44
					2	−6.96	3.51×10^{-12}				
rs3817198	11:1865582	LSP1	T/C	0.34	1	4.17	3.06×10^{-5}	0.14 (0.03)	6.47	9.67×10^{-11}	0.99
					2	4.97	6.63×10^{-7}				
rs703556	12:101536024	IGF1	A/G	0.02	1	−5.13	2.91×10^{-7}	−0.41 (0.08)	−6.27	3.74×10^{-10}	0.90
					2	−3.85	0.0001				
rs7289126	22:36958252	TMEM184B	C/A	0.40	1	−4.30	1.74×10^{-5}	−0.11 (0.02)	−5.55	2.80×10^{-8}	0.99
					2	−3.63	0.0003				
rs17001868	22:39108177	SGSM3, MKL1	A/C	0.08	1	−4.42	9.99×10^{-6}	−0.18 (0.03)	−7.33	2.29×10^{-13}	0.74
					2	−5.87	4.35×10^{-9}				
SNPs associated with mammographic non-dense area											
rs7816345	8:36965267	N/A	C/T	0.18	1	−7.37	1.77×10^{-13}	−0.24 (0.03)	−9.96	2.40×10^{-23}	0.09
					2	−6.77	1.30×10^{-11}				
SNPs associated with mammographic percent density											
rs186749	5:122482204	PRDM6	G/A	0.28	1	4.98	6.26×10^{-7}	0.10 (0.02)	5.96	2.52×10^{-9}	0.43
					2	3.31	0.0009				
rs7816345	8:36965267	N/A	C/T	0.18	1	2.93	0.003	0.08 (0.02)	5.46	4.68×10^{-8}	0.74
					2	4.79	1.69×10^{-6}				
rs7289126	22:36958252	TMEM184B	C/A	0.40	1	−4.75	1.99×10^{-6}	−0.10 (0.02)	−5.86	4.68×10^{-9}	0.82

Lead SNP	Chr: Position[1]	Gene[2]	Alleles[3]	MAF[4]	Stage	Z-Score	P	β[5] (se)	Z-Score	P	P (Het)[6]
									Stage 1+2		
					2	−3.54	0.0004				

[1] HG version 18,

[2] Includes nearby genes,

[3] Major allele (Reference allele)/Minor allele (Effect allele),

[4] Minor Allele Frequencies (MAF) as in the 1000 Genomes project.

[5] β and standard error estimates were obtained using fixed-effects meta-analysis of cross-sectional studies (i.e. studies that analyzed density phenotypes as a quantitative trait) in unrelated individuals. Mammographic density phenotypes are square-root transformed.

[6] P-value for heterogeneity between studies.

Table 2

Breast cancer associations for mammographic density SNPs based on a meta-analysis of 62,533 breast cancer cases and 60,976 controls.

Mammographic density phenotype	SNP	Chr	Gene	Alleles[1]	Z-score, mammographic density association (density phenotype[2])	Breast Cancer Association OR (95% CI)	P[3] Breast Cancer Association
Dense Area	rs10034692	4	AREG	A/G	−6.35 (DA)	0.99 (0.97–1.01)	0.31
	rs12665607	6	ESR1	T/A	5.64 (DA)	1.20 (1.16–1.23)	1.48×10^{-30}
	rs3817198	11	LSP1	T/C	6.47 (DA)	1.07 (1.05–1.09)	2.09×10^{-13}
	rs703556	12	IGF1	A/G	−6.27 (DA)	0.94 (0.90–0.99)	0.02
	rs17001868	22	SGSM3/MKL1	A/C	−7.33 (DA)	1.10 (1.08–1.13)	1.19×10^{-15}
Percent Density	rs186749	5	PRDM6	G/A	5.96 (PD)	1.02 (1.01–1.04)	0.009
Dense Area and Percent Density	rs10995190	10	ZNF365	G/A	−8.26 (DA)	0.86 (0.84–0.88)	1.50×10^{-37}
	rs7289126	22	TMEM184B	C/A	−5.55 (DA), −5.86 (PD)	0.98 (0.97–1.00)	0.03
Non-dense Area and Percent Density	rs7816345	8	N/A	C/T	−9.96 (NDA), 5.46 (PD)	0.94 (0.92–0.96)	2.18×10^{-8}

[1] Reference allele/Effect allele

[2] DA=dense area; PD=percent density; NDA=Non-dense area

[3] P-value based on a fixed effects meta-analysis. Estimates and standard errors for each breast cancer study were calculated using logistic regression.