Identification of a novel percent mammographic density locus at 12q24

Kristen N. Stevens1, Sara Lindstrom7,8,9, Christopher G. Scott1, Deborah Thompson10,11, Thomas A. Sellers13, Xianshu Wang2, Alice Wang1, Elizabeth Atkinson1, David N. Rider1, Jeanette E. Eckel-Passow1, Jajini S. Varghese10,11, Tina Audney10,11, Judith Brown10,11, Jean Leyland10,11, Robert N. Luben10,11, Ruth M.L. Warren12, Ruth J.F. Loos14, Nicholas J. Wareham14, Jingmei Li15, Per Hall16, Jianjun Liu15, Louise Eriksson16, Kamila Czene16, Janet E. Olson1, V. Shane Pankratz1, Zachary Fredericksen1, Robert B. Diasio3, Adam M. Lee3, John A. Heit4, Mariza deAndrade1, Ellen L. Goode1, Robert A. Vierkant1, Julie M. Cunningham2, Sebastian M. Armasu1, Richard Weinshilboum5, Brooke L. Fridley1, Anthony Batzler1, James N. Ingle6, Norman F. Boyd17, Andrew D. Paterson18,19, Johanna Rommens19, Lisa J. Martin17, John L. Hopper20, Melissa C. Southey21, Jennifer Stone20, Carmel Apicella20, Peter Kraft2,8,9, Susan E. Hankinson7,22, Aditi Hazra22, David J. Hunter7,8,22, Douglas F. Easton10,11, Fergus J. Couch2, Rulla M. Tamimi7,22 and Celine M. Vachon1,*

1Department of Health Sciences Research, 2Department of Laboratory Medicine and Pathology, 3Department of Molecular Pharmacology and Experimental Therapeutics, 4Division of Cardiovascular Disease, Department of Medicine, 5Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics and 6Department of Oncology, Mayo Clinic, Rochester, MN 55905, USA, 7Department of Epidemiology, 8Program in Molecular and Genetic Epidemiology and 9Department of Biostatistics, Harvard School Of Public Health, Boston, MA 02138, USA, 10Department of Public Health and Primary Care, 11Department of Oncology and 12Department of Radiology, University of Cambridge, Addenbrooke’s NHS Foundation Trust, Cambridge CB2 1TN, UK, 13Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL 33612, USA, 14MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge CB2 1TN, UK, 15Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore, 16Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm SE-171 77, Sweden, 17Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute, Toronto, Ontario, Canada M5G 2M9, 18Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada MSS 1A1, 19Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada MSG 1X8, 20Centre for Molecular, Environmental, Genetic and Analytic Epidemiology School of Population Health, The University of Melbourne, Melbourne 3010, Australia, 21Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne 3010, Australia and 22Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA 02138, USA

Received January 16, 2012; Revised March 29, 2012; Accepted April 17, 2012

Percent mammographic density adjusted for age and body mass index (BMI) is one of the strongest risk factors for breast cancer and has a heritable component that remains largely unidentified. We performed a three-stage genome-wide association study (GWAS) of percent mammographic density to identify novel genetic loci associated with this trait. In stage 1, we combined three GWASs of percent density comprised of 1241

*To whom correspondence should be addressed at: Department of Health Sciences Research, Mayo Clinic, Charlton 6-239, 200 First St. SW, Rochester, MN 55905, USA. Tel: +1 5072849901; Fax: +1 5072662478; Email: vachon.celine@mayo.edu

© The Author 2012. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
women from studies at the Mayo Clinic and identified the top 48 loci (99 single nucleotide polymorphisms). We attempted replication of these loci in 7018 women from seven additional studies (stage 2). The meta-analysis of stage 1 and 2 data identified a novel locus, rs1265507 on 12q24, associated with percent density, adjusting for age and BMI ($P = 4.43 \times 10^{-8}$). We refined the 12q24 locus with 459 additional variants (stage 3) in a combined analysis of all three stages ($n = 10$ 377) and confirmed that rs1265507 has the strongest association in the 12q24 region ($P = 1.03 \times 10^{-8}$). Rs1265507 is located between the genes *TBX5* and *TBX3*, which are members of the phylogenetically conserved T-box gene family and encode transcription factors involved in developmental regulation. Understanding the mechanism underlying this association will provide insight into the genetics of breast tissue composition.

**INTRODUCTION**

Percent mammographic density (PD) is an estimate of the proportion of stromal and epithelial breast tissues quantified from a mammogram image. Adjusted for age and body mass index (BMI), PD is one of the strongest risk factors for breast cancer, and women in the highest quartile of density have a 3–5-fold increased risk compared with women in the lowest quartile (1). Other known breast cancer risk factors, primarily BMI and age, are estimated to account for up to 20–30% of the variation in mammographic density (2,3). Twin and family studies suggest that up to two-thirds of the remaining variability are accounted for by genetic factors (4,5). Indeed, a recent genome-wide association study (GWAS) of PD adjusted for age and BMI identified a genome-wide significant association with a common genetic variant in the *ZNF365* breast cancer susceptibility locus; there was evidence that the association between *ZNF365* and breast cancer risk may be mediated through PD (6). However, these single nucleotide polymorphisms (SNPs) are estimated to account for only 0.5–1.0% of the variance in adjusted PD, leaving much of the genetic contribution to this trait unexplained. We performed a three-stage GWAS to identify additional common genetic variation associated with the PD measure that predicts breast cancer risk.

**RESULTS**

We first evaluated associations between genome-wide SNPs and PD in a pooled analysis of three independent studies at the Mayo Clinic for which phenotype, risk factor and genome-wide scan data were available (stage 1). These three studies collectively totaled 1241 women without breast cancer, involving 571 members of the Minnesota Breast Cancer Family Study (MBCFS), 317 controls from a case–control study of venous thromboembolism (Mayo VTE) and 363 controls from the Mayo Ovarian Cancer Case Control Study (MAY) (Supplementary Material, Tables S1 and S2). We analyzed SNPs that were either directly genotyped on the Illumina 660W Quad or 610 Quad arrays, or imputed using Phase II HapMap data on subjects of European ancestry (CEU) with an imputation quality score above 0.6 (Supplementary Material, Table S2). Demographic characteristics of the participants were similar across studies; all women were of European ancestry and the majority (74%) were post-menopausal at the time of mammogram (Supplementary Material, Tables S1 and S3). PD was measured by the same reader using the Cumulus software (7).

Overall, there was no evidence for genomic inflation among the directly genotyped ($\lambda = 1.01$) or imputed ($\lambda = 1.00$) SNPs in the analysis for association with PD adjusted for age and BMI (Supplementary Material, Figs S1 and S2). No SNPs achieved genome-wide significance ($P < 5 \times 10^{-8}$) in the stage 1 analysis. We selected the top 99 SNPs representing 48 loci from the combined GWAS results, taking into account both overall significance and consistency of effect estimates across the three studies (Materials and Methods).

To replicate these findings, in stage 2, we evaluated the association between these 99 variants and PD adjusted for age and BMI in seven additional studies of PD in the Markers of Density (MODE) consortium comprised of 7018 women of European ancestry (stage 2: EPIC-Norfolk $n = 1142$ controls; NHS Phase 1 $n = 1590$ breast cancer cases and controls; NHS Phase 2 $n = 778$ breast cancer cases and controls; SASBAC $n = 526$ breast cancer cases; SASBAC $n = 742$ controls; Toronto/Melbourne $n = 316$ controls; SIBS $n = 1145$ controls) (Supplementary Material, Tables S1, S3 and S4). Of the 99 SNPs, rs1265507 on chromosome12q24 had the strongest association with PD in stage 2 ($P = 2.74 \times 10^{-7}$).

We then performed a meta-analysis of stage 1 and stage 2 data for these 99 SNPs ($n = 8269$). As described previously (6), we combined $P$-values and the direction of association weighted by the square root of the sample size and study-specific inflation factors, rather than combining effect estimates, due to differences in study design. In the combined analysis of studies from stages 1 and 2 (Supplementary Material, Table S5), a single SNP, rs1265507, displayed a genome-wide significant association with PD adjusting for age and BMI ($P = 4.43 \times 10^{-8}$). The T allele of rs1265507 was inversely associated with PD in 9 of 10 studies (Table 1). There was no evidence for heterogeneity of the association across studies ($P$ heterogeneity = 0.12).

To determine whether other SNPs in the region also demonstrated strong evidence for association with PD and to refine the region of association with PD on 12q24 particularly with respect to the nearby genes *TBX5* and *TBX3*, we evaluated 459 additional variants at the 12q24 locus spanning a 407 kb region. These SNPs were selected to cover the nearby genes *TBX5* and *TBX3*, including all SNPs that were either directly genotyped on the Illumina 660W Quad or 610 Quad arrays or imputed using Phase II HapMap data. Although these variants were previously examined in the stage 1 GWAS.
variant (2 and 3, rs1265507 remained the most significantly associated Tables S1, S2 and S4). In the combined analysis of stages 1, cancer cases, genome-wide significant associations with breast cancer (8), 12q24 over 250 kb distal to rs1265507 have recently shown attenuate the significance of the association (of breast cancer cases from the rs1265507 analysis slightly
2 (6.61 × 10⁻⁸; rs2551389 P = 3.54 × 10⁻⁷) (Supplementary Material, Tables S9 and S10). Across studies with genotype
data for rs1265507 (excluding the Toronto/Melbourne study), rs1265507 was associated with a mean change of
× 10⁻⁸. The remaining 458 variants are shown as circles, colored by the degree of linkage disequilibrium (R²) between each SNP and rs1265507. The continuous blue line represents the recombination rate (cM/Mb).
and few were selected for the replication, it is possible that some variants had stronger effects in the stage 2 or 3 studies. As such, we performed a meta-analysis for these 459 SNPs using data from 9126 samples from the three stage 1 GWAS studies, the seven stage 2 MODE studies and data from two additional studies (stage 3: Mayo Clinic PGRN n = 213 breast cancer cases; MCBCS n = 892 breast cancer cases, n = 1,003 controls) (Supplementary Material, Tables S1, S2 and S4). In the combined analysis of stages 1, 2 and 3, rs1265507 remained the most significantly associated variant (P = 1.03 × 10⁻⁸) with no evidence for heterogeneity across studies (P heterogeneity = 0.17) (Table 1). Exclusion of breast cancer cases from the rs1265507 analysis slightly attenuated the significance of the association (P = 2.98 × 10⁻⁸) (Supplementary Material, Table S6). An additional 39 variants in moderate to strong LD with rs1265507 in a 33.2 kb region surrounding rs1265507 were also associated with PD (P < 5 × 10⁻⁸) (Fig. 1, Supplementary Material, Table S7). Of these, rs2551389 was in perfect LD with rs1265507 (R² = 1) and also achieved genome-wide significance (P = 2.67 × 10⁻⁸) (Supplementary Material, Table S8). We also found that these two 12q24 SNPs were associated with absolute dense area, with some attenuation of significance compared with the PD result [rs1265507 P = 6.61 × 10⁻⁸; rs2551389 P = 3.54 × 10⁻⁷] (Supplementary Material, Tables S9 and S10). Across studies with genotype data for rs1265507 (excluding the Toronto/Melbourne study), rs1265507 was associated with a mean change of −1.03% in PD per copy of the minor allele. Based on this estimate, rs1265507 would explain at most 1.3% of the variance in percent mammographic density.
Given that age- and BMI-adjusted mammographic density is a strong risk factor for breast cancer and that variants from 12q24 over 250 kb distal to rs1265507 have recently shown genome-wide significant associations with breast cancer (8), we evaluated whether rs1265507 was associated with the risk of breast cancer in MCBCS, SASBAC, the Cancer Genetic Markers of Susceptibility (CGEMS) study (9) and the UK2 GWAS (10) (Supplementary Methods, Supplementary Material, Table S11). No significant evidence for association with breast cancer was observed [combined odds ratio (OR) = 0.99, 95% confidence interval (CI) 0.95–1.04, P = 0.67] (Supplementary Material, Table S12), suggesting that any causal variants underlying the association between rs1265507 and adjusted PD are not strong independent risk factors for breast cancer. However, it remains possible that variants underlying the independent mammographic density and breast cancer associations have common biological effects in this region.

Table 1. Rs1265507 association with percent mammographic density

<table>
<thead>
<tr>
<th>Study</th>
<th>Genotyped or imputeda</th>
<th>N</th>
<th>MAFb</th>
<th>b</th>
<th>95% CI</th>
<th>P-value</th>
<th>P-het</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1: Mayo Clinic GWAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBCFS Genotyped</td>
<td></td>
<td>571</td>
<td>0.46</td>
<td>−0.31</td>
<td>(−0.46, −0.15)</td>
<td>1.60 × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>Mayo VTE Genotyped</td>
<td></td>
<td>317</td>
<td>0.47</td>
<td>−0.38</td>
<td>(−0.60, −0.15)</td>
<td>1.08 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>MAY Genotyped</td>
<td></td>
<td>363</td>
<td>0.44</td>
<td>0.0098</td>
<td>(−0.18, 0.20)</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Stage 1 combined</td>
<td>1241c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.31 × 10⁻⁵</td>
<td></td>
</tr>
<tr>
<td>Stage 2: replication (MODE studies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPIC Imputed (0.99)</td>
<td>1129</td>
<td>0.46</td>
<td>−0.073</td>
<td>(−0.15, −0.001)</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHS Phase 1 Genotyped</td>
<td>1590</td>
<td>0.47</td>
<td>−0.093</td>
<td>(−0.21, 0.023)</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHS Phase 2 Genotyped</td>
<td>778</td>
<td>0.45</td>
<td>−0.14</td>
<td>(−0.30, 0.020)</td>
<td>0.092</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SASBAC cases Genotyped</td>
<td>525</td>
<td>0.45</td>
<td>−0.15</td>
<td>(−0.35, 0.040)</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SASBAC controls</td>
<td>742</td>
<td>0.46</td>
<td>−0.14</td>
<td>(−0.30, 0.018)</td>
<td>0.087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toronto/Melbourne Genotyped</td>
<td>1109d</td>
<td>0.49</td>
<td>−0.67</td>
<td>(−1.77, 0.43)</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIBS Imputed (0.96)</td>
<td>1145</td>
<td>0.46</td>
<td>−0.075</td>
<td>(−0.20, 0.046)</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>7018</td>
<td></td>
<td>Z-score = −4.19</td>
<td>2.74 × 10⁻⁵</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1 and 2 combined</td>
<td>8269</td>
<td></td>
<td>Z-score = −5.47</td>
<td>4.43 × 10⁻⁸</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 3: additional chromosome 12 fine-mapping studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayo PGRN Genotyped</td>
<td>213</td>
<td>0.47</td>
<td>−0.23</td>
<td>(−0.58, 0.11)</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCBCS Genotyped</td>
<td>1895</td>
<td>0.47</td>
<td>−0.067</td>
<td>(−0.15, 0.019)</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages 1, 2 and 3 combined</td>
<td>10,377</td>
<td></td>
<td>Z-score = −5.73</td>
<td>1.03 × 10⁻⁸</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aIf imputed, imputation quality score indicated in parentheses.
bMinor allele frequency (T allele).
cTen samples excluded in the combined Mayo Clinic GWAS, see Supplementary Methods.
dRepresents the weighted Toronto/Melbourne sample size (n = 316, scale factor = 3.51).

Figure 1. Association between 12q24.21 variants (n = 459) and percent mammographic density. The association between 459 variants from the combined 12q24 analyses in stages 1, 2 and 3 are shown. The most significant SNP (rs1265507) is shown as the purple diamond (rs1265507) is shown as the purple diamond (P = 5.47 × 10⁻⁸). The remaining 458 variants are shown as circles, colored by the degree of linkage disequilibrium (R²) between each SNP and rs1265507. The continuous blue line represents the recombination rate (cM/Mb).
DISCUSSION

Through a three-stage GWAS, we describe a novel genome-wide significant association between rs1265507 at 12q24 and PD adjusting for age and BMI ($P = 1.03 \times 10^{-8}$). We further demonstrate that this SNP had the strongest association of those examined in the 12q24 region, and was located in a cluster of SNPs surrounding rs1265507 also associated with PD ($P < 5 \times 10^{-4}$). This is only the second genetic locus identified for age- and BMI-adjusted PD to date.

Rs1265507 is located ~22 kb from the 5′ untranslated region (UTR) of the T-box 5 (TBX5) gene and ~240 kb from the 3′ UTR of the T-box 3 (TBX3) gene. TBX5 and TBX3 are both members of the phylogenetically conserved T-box gene family, which encode transcription factors involved in developmental regulation. Interestingly, the cluster of 39 SNPs most strongly associated with age- and BMI-adjusted PD surrounding rs1265507 appears to be located in a region flanked by the coding region of TBX3 and a recombination hotspot (recombination rate ~60 cM/Mb) (Fig. 1). This raises the possibility that the underlying causal variant(s) influence the regulation of the TBX3 promoter. While rs1265507 was not associated with expression of either TBX3 or TBX3 as reported in two databases of expression quantitative trait loci (eQTL), the Wellcome Trust Genevar database and the University of Chicago eQTL browser (11,12), these databases did not include expression data from normal breast epithelium or stromal tissue. Mutations in TBX5 are associated with Holt-Oram syndrome, a developmental disorder affecting the heart and upper limbs (13). TBX5 has also recently been implicated as a tumor suppressor gene that is epigenetically inactivated in colon cancer (14). In addition, TBX3 has been implicated in embryonic development of the mammary gland (15,16), and mutations in this gene have been linked to ulnar-mammary syndrome (17). TBX3 is also overexpressed in breast tumors (18), and a recent study using an inducible transgenic mouse model showed that TBX3 overexpression induces mammary gland hyperplasia and increases mammary stem-like cells (19).

We recognize that the design of participant studies used for this GWAS varied, including the use of different mammography views and genotyping platforms. However, the consistency of association across studies supports the validity of this finding. Additionally, a strength of this study is that a single reader estimated PD for the stage 1 studies, thereby reducing noise in our density phenotype and increasing our power to detect genetic associations with PD. Also, our increased sample size from the original MODE GWAS (6) (8269 in stages 1–2 compared with 4877 in MODE), coupled with a more homogenous density phenotype in stage 1, resulted in the identification and confirmation of the 12q24 locus. The use of imputed genotype data in our replication and fine-mapping analyses may have affected the estimated strength of association for all variants, and precluded our ability to detect genome-wide significance for some variants; however, GWAS studies of mammographic density measures with larger sample sizes will help to clarify these associations and to identify additional common variants associated with this trait.

This is only the second genetic locus identified for age- and BMI-adjusted PD, and unlike the ZNF365 locus (6), does not appear to be associated with breast cancer risk. As we learned from this association between PD and the ZNF365 locus, the identification of genetic associations with PD may correspond to known or novel breast cancer risk loci due to the strong relationship between mammographic density and breast cancer risk. However, although age- and BMI-adjusted percent mammographic density is a predictor of breast cancer risk, the majority of women with high breast density never develop breast cancer. It may be important to identify genetic factors that are not related to breast cancer risk to aid in discriminating between biological pathways that are crucial to breast tumorigenesis in dense breast tissue and those that are not. This, in turn, may help explain why only a portion of women with high breast density eventually develop breast cancer. In addition, newly identified genetic variation associated with density could lead to improved methods for visualization of cancers in dense tissue. This could include the development of new pharmaceutical agents for decreasing dense tissue in a woman to allow better screening for breast cancer or even new technologies relating to better imaging of dense tissue.

This study provides additional evidence that common genetic variation contributes to breast tissue composition. Resequencing, fine-mapping and functional studies of the 12q24 locus in additional studies of mammographic density are necessary to understand the causal variants and mechanism underlying this novel association.

MATERIALS AND METHODS

Mayo Clinic GWAS studies

Three independent studies at the Mayo Clinic (Rochester, MN, USA) contributed genotype and phenotype data to the stage 1 GWAS and meta-analyses. Women of white European ancestry with GWAS and density data were included from the Minnesota Breast Cancer Family Study (MBCFS $n = 571$ controls), the Mayo Venous Thromboembolism Study (Mayo VTE $n = 317$ controls) and the Mayo Clinic Ovarian Case Control Study (MAY $n = 363$ controls). These studies are described in detail in Supplementary Material.

MODE studies

Seven independent studies from the Markers of Density (MODE) consortium contributed summary statistics from analyses of associations between stage 2 ($n = 99$) and stage 3 ($n = 459$) SNPs and adjusted PD. These studies were comprised of EPIC-Norfolk ($n = 1142$ controls), the Nurses’ Health Study (NHS Phase 1 $n = 1590$ breast cancer cases and controls; Phase 2 $n = 778$ breast cancer cases and controls), the Singapore and Sweden Breast Cancer Study (SASBAC $n = 526$ breast cancer cases; $n = 742$ controls), the Toronto/Melbourne study ($n = 316$ controls) and the Sisters in Breast Screening study (SIBS $n = 1145$ controls). All MODE studies were comprised of women of self-reported European ancestry. These studies are also described in detail in Supplementary Material.
Additional chromosome 12 fine-mapping studies

Two additional, independent studies from the Mayo Clinic contributed data to the Stage 3 meta-analysis of the 459 SNPs on chromosome 12q24. These included the Mayo Pharmacogenetics Research Network Aromatase Inhibitor (AI) Study (Mayo PGRN n = 213 breast cancer cases) and the Mayo Clinic Breast Cancer Study (MCBCS n = 892 breast cancer cases, n = 1003 controls). The Mayo PGRN and MCBCS contained only women of white European ancestry and are also described in detail in Supplementary Material.

Percent mammographic density measurement

Mammogram collection and estimation of percent density are described for each study in Supplementary Material. Percent density was measured from mammograms using the craniocaudal view, the mediolateral-oblique view or the average of the two views using the Cumulus software (7) (Supplementary Material, Table S2).

Stage 1: GWAS genotyping, imputation and analyses

MBCFS and MAY samples were genotyped at the Mayo Clinic Genotyping Shared Resource on the Illumina Human-Hap 660W Quad and Illumina 610K arrays, respectively. Mayo VTE genotyping was performed as part of a larger GWAS of VTE using the Illumina 660W Quad array at the Center for Inherited Disease Research at Johns Hopkins University. Prior to sample exclusions, there were 597 MBCFS, 427 MAY and 638 Mayo VTE samples for a total of 1662 genotyped samples available for the pooled GWAS analysis. Samples were excluded for the following reasons: call rates < 98% (n = 31), duplicates across the studies (n = 21) and gender (n = 4) (Supplementary Material, Table S2). Genotype data for the remaining 1625 women were combined using Plink. All failed SNPs, monomorphic SNPs and SNPs with call rates < 95% were excluded, resulting in 521 571 SNPs in the combined Mayo GWAS data set. MACH version 1.0.16 was used to impute autosomal SNPs for the remaining 1625 samples using the Hapmap II build 36 CEU reference population. A total of 2,510,880 SNPs with $R^2 > 0.6$ or MAF > 1% were retained for the analysis. After imputation, we then excluded samples that were possibly related (n = 3), non-European (n = 1), VTE cases (n = 167) and those that did not have complete phenotype and covariate data (n = 190), leaving a total of 1241 samples for analysis.

The final analytic data set contained 1241 women with complete mammographic density, covariate and genotype data: 571 MBCFS, 313 Mayo VTE and 357 MAY. Information on 1942 additional MBCFS family members without phenotype data was used to inform family structure for analyses. Linear mixed-effects models were fit to test the additive association of genotype with square root percent density after adjusting for age, BMI, menopausal status and study while also accounting for genetic relatedness among individuals through the incorporation of a polygenic random effect into the model. We assessed the degree to which population stratification influenced our results by adjusting for the first two principal components (PC) from a PC analysis utilizing ancestry informative markers. Inclusion of these PCs did not affect the results of the analysis and suggested that population stratification is minimal. These analyses were performed in the multilocus package in R (http://cran.r-project.org/).

SNP selection for replication

We selected SNPs from stage 1 for replication using the following method. First, we considered overall significance of SNP associations in the pooled stage 1 analysis, and selected SNPs in order of $P$-value rank. We retained these SNPs if they were associated with PD in the same direction in at least two of the three stage 1 studies; had a HWE $P > 0.001$; and if imputed (rather than genotyped), had an imputation quality score ($R^2$) > 0.8. Additional SNPs from the same locus were included for replication in the following situations: (i) additional SNPs in the region had $P$-values < 0.01 or (ii) if the most significant SNPs in the locus was imputed, the genotyped SNP with the lowest $P$-value and highest LD was also included for replication. SNPs from the same locus were not filtered to be in low pair-wise LD with each other.

Stages 2 and 3: genotyping, imputation and analyses

The individual MODE studies were genotyped on multiple platforms and used various quality control assessments (Supplementary Material, Table S4). All MODE studies used identical-by-state or identical-by-descent measurements to identify and exclude individuals with unexpected relatedness. Since the SASBAC genotyped their breast cancer cases and controls on separate platforms, they were imputed and analyzed separately. All studies, other than Mayo PGRN and MCBCS, performed imputation to obtain a total of ~2.5 million autosomal genotypes using HapMap Phase II CEU samples as reference. To reassure high-quality data for all studies, we excluded all SNPs with a minor allele frequency below 0.01 or an imputation quality score below 0.6 [as defined by the RSQR_HAT value in MACH (20,21), the PROPER_INFO in IMPUTE (22,23) and the information content (INFO) measure in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/) (24)]. Study-specific summary estimates from linear regression analyses of square root PD (see Supplementary Material, Table S4) were obtained for the 99 and 459 SNPs of interest in stages 2 and 3, respectively. All studies were adjusted for age and BMI; NHS, SASBAC and the Toronto/Melbourne study adjusted for population stratification using principal components analysis. Additional study-specific adjustments are described in Supplementary Material, Table S4.

Mayo PGRN subjects who provided blood samples as a source of DNA (n = 835) were genotyped using the Illumina 610K array, performed at RIKEN, with 830 cases passing genotype quality control. One hundred four genotyped SNPs of the 459 SNPs of interest in the 12q24 region were analyzed for association with the square root of PD adjusted for age, BMI, menopausal status and breast cancer case–control status. Genotyping of the rs1265507 SNP in MCBCS was performed using a TaqMan assay at the Mayo Clinic Genotyping Shared Resource. A total of 95 CEPH controls on a Coriell test
plate, HAPMAPPT01, were also typed simultaneously to establish genotyping accuracy. Genotyping of the rs1265507 SNP displayed high SNP call rates (>99% for cases, controls and overall), high concordant rate (100%) and no deviation from Hardy–Weinberg equilibrium among controls (P = 0.52). Rs1265507 was analyzed for association with square root of PD adjusting for age, BMI and menopausal status.

**Meta-analyses**

All meta-analyses were based on summary statistics from contributing studies. The study-specific summary statistics obtained were sample size, alleles tested, minor allele frequency, effect estimate (β), standard error and P-value. For each SNP, we combined study-specific P-values and direction of association using the METAL software (25). Weights were proportional to study-specific sample size and P-values were adjusted with genomic inflation factors, where relevant. To account for the extreme sampling scheme in the Toronto/Melbourne study, we up-weighted this study with a scale factor of 3.51. The scale factor was determined as described by Lindstrom et al. (6). We used Cochran’s Q statistic to test for heterogeneity across studies.

The meta-analysis of the top 48 loci from the Mayo Clinic GWAS for stage 1 and 2 combined was based on summary statistics for 99 SNPs (Supplementary Material, Table S4) from 10 studies in a total of up to 8283 Caucasian women. Studies included in this analysis were the three Mayo Clinic GWAS and the seven Markers of Density (MODE) studies (SASBAC cases and controls separately, EPIC-Norfolk, NHS phases 1 and 2 separately, Toronto/Melbourne, SIBS). Analyses of each study-specific MODE GWAS from which these estimates were obtained have been described previously (6). Briefly, similar to above, the square root of PD was used as the outcome, adjusting for age, BMI, menopausal status and case-control status, when applicable.

Meta-analysis of 459 chromosomes 12q24.21 fine-mapping SNPs was based on summary statistics from 11 studies in a total of up to 10 377 Caucasian women. Studies included in this analysis were the three Mayo Clinic GWAS, the seven MODE studies, Mayo PGRN and MCBCS. The Mayo PGRN study contributed data only for those 104 SNPs which were directly genotyped on the Illumina 610K platform. The final rs1265507 combined analyses of all studies (stages 1–3) were performed using summary statistics from the three Mayo Clinic GWAS studies, seven MODE studies, Mayo PGRN and MCBCS. Figure 1 was generated using the LocuzZoom tool (26).

**Association with breast cancer**

The association between rs1265507 and risk of breast cancer was examined in MCBCS, SASBAC, the Cancer Genetic Markers of Susceptibility study (CGEMS) (9) and a GWAS of breast cancer in the UK (UK2 GWAS) (10). The odds ratio, 95% confidence interval and P-value for the rs1265507 association with breast cancer in CGEMS were obtained using publically available data. For MCBCS, SASBAC and the UK2 GWAS, analyses were conducted using logistic regression. MCBCS adjusted for age, 1/BMI, menopausal status and SASBAC adjusted for three principal components, age, BMI and menopausal status. The UK2 GWAS adjusted for two principal components. A combined estimate was calculated using the metagen function in the ‘meta’ R package (http://cran.r-project.org/).

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

**FUNDING**

This work and the MBCFS study were supported by National Institutes of Health grant R01 CA128931. The Mayo VTE was supported by National Heart, Lung and Blood Institute grant HL83131; National Human Genome Research Institute grant HG04735; National Cancer Institute grant CA92153; and Centers for Disease Control grant DD000235. MAY was supported by National Institutes of Health grants R01 CA122443, R01 CA114343 and P50 CA136393. The EPIC-Norfolk study was funded by the 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. NHS was supported by Public Health Service Grants CA131332, CA087969, CA049449, CA089393 from the National Cancer Institute, National Institutes of Health, National Institutes of Health, Department of Health and Human Services and Breast Cancer Research Fund. The NHS Phase I breast cancer cases and controls were genotyped with support from the National Cancer Institute’s Cancer Markers of Susceptibility (CGEMS) initiative. SASBAC was supported by National Institutes of Health grant R01 CA58427; Märit and Hans Rausing’s Initiative Against Breast Cancer; the W81XWH-05-1-0314 Innovator Award; US Department of Defense Breast Cancer Research Program; Office of the Congressionally Directed Medical Research Programs; and the Agency for Science, Technology and Research (A*STAR). Genotyping in the TORONTO/MELBOURNE subjects was supported by the Campbell Family Institute for Breast Cancer Research. Support was also provided by the Ontario Ministry of Health and Long Term Care. The SIBS study was supported by Cancer Research UK programme grant C1287/A10118 and Cancer Research UK project grants C1287 and C8459. D.F.E. is a Principal Research Fellow of Cancer Research UK. MCBCS was supported by Mayo Clinic Breast Cancer SPORE P50 CA116201 and National Institutes of Health grant R01 CA122340.

MAYO PGRN was supported by Mayo Clinic Pharmacogenomics Research Network Center U19 GM61388 and the Mayo Clinic Breast Cancer SPORE P50 CA116201.
REFERENCES


