

## Mammographic Breast Density and Breast Cancer: Evidence of a Shared Genetic Basis

Jajini S. Varghese<sup>1</sup>, Deborah J. Thompson<sup>1</sup>, Kyriaki Michailidou<sup>1</sup>, Sara Lindström<sup>4</sup>, Clare Turnbull<sup>6</sup>, Judith Brown<sup>1</sup>, Jean Leyland<sup>1</sup>, Ruth M.L. Warren<sup>2</sup>, Robert N. Luben<sup>1</sup>, Ruth J. Loos<sup>3</sup>, Nicholas J. Wareham<sup>3</sup>, Johanna Rommens<sup>7</sup>, Andrew D. Paterson<sup>7,8</sup>, Lisa J. Martin<sup>9</sup>, Celine M. Vachon<sup>10</sup>, Christopher G. Scott<sup>10</sup>, Elizabeth J. Atkinson<sup>10</sup>, Fergus J. Couch<sup>11</sup>, Carmel Apicella<sup>12</sup>, Melissa C. Southey<sup>13</sup>, Jennifer Stone<sup>12</sup>, Jingmei Li<sup>14,15</sup>, Louise Eriksson<sup>14</sup>, Kamila Czene<sup>14</sup>, Norman F. Boyd<sup>9</sup>, Per Hall<sup>14</sup>, John L. Hopper<sup>12</sup>, and Rulla M. Tamimi<sup>4,5</sup> for the MODE Consortium, Nazneen Rahman<sup>6</sup>, and Douglas F. Easton<sup>1</sup>

### Abstract

Percent mammographic breast density (PMD) is a strong heritable risk factor for breast cancer. However, the pathways through which this risk is mediated are still unclear. To explore whether PMD and breast cancer have a shared genetic basis, we identified genetic variants most strongly associated with PMD in a published meta-analysis of five genome-wide association studies (GWAS) and used these to construct risk scores for 3,628 breast cancer cases and 5,190 controls from the UK2 GWAS of breast cancer. The signed per-allele effect estimates of single-nucleotide polymorphisms (SNP) were multiplied with the respective allele counts in the individual and summed over all SNPs to derive the risk score for an individual. These scores were included as the exposure variable in a logistic regression model with breast cancer case-control status as the outcome. This analysis was repeated using 10 different cutoff points for the most significant density SNPs (1%–10% representing 5,222–50,899 SNPs). Permutation analysis was also conducted across all 10 cutoff points. The association between risk score and breast cancer was significant for all cutoff points from 3% to 10% of top density SNPs, being most significant for the 6% (2-sided  $P = 0.002$ ) to 10% ( $P = 0.001$ ) cutoff points (overall permutation  $P = 0.003$ ). Women in the top 10% of the risk score distribution had a 31% increased risk of breast cancer [OR = 1.31; 95% confidence interval (CI), 1.08–1.59] compared with women in the bottom 10%. Together, our results show that PMD and breast cancer have a shared genetic basis that is mediated through a large number of common variants. *Cancer Res*; 72(6); 1478–84. ©2012 AACR.

**Authors' Affiliations:** <sup>1</sup>Department of Public Health and Primary Care, <sup>2</sup>Department of Radiology, Addenbrooke's NHS Foundation Trust, <sup>3</sup>MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom; <sup>4</sup>Department of Epidemiology, Harvard School of Public Health; <sup>5</sup>Channing Laboratory, Brigham and Women's Hospital, Boston, Massachusetts; <sup>6</sup>Section of Cancer Genetics, The Institute of Cancer Research, Sutton, Surrey, United Kingdom; <sup>7</sup>Program in Genetics and Genome Biology, The Hospital for Sick Children; <sup>8</sup>Lana School of Public Health, University of Toronto; <sup>9</sup>Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute, Toronto, Ontario, Canada; Departments of <sup>10</sup>Health Sciences Research and <sup>11</sup>Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota; <sup>12</sup>Centre for M.E.G.A. Epidemiology, Melbourne School of Population Health, <sup>13</sup>Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia; <sup>14</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; and <sup>15</sup>Human Genetics, Genome Institute of Singapore, Singapore, Singapore

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Corresponding Author:** Deborah J. Thompson, Centre for Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratories, Worts Causeway, Cambridge CB1 8RN, United Kingdom. Phone: 44-1223-740164; Fax: 44-1223-740159; E-mail: [deborah.thompson@phpc.cam.ac.uk](mailto:deborah.thompson@phpc.cam.ac.uk)

doi: 10.1158/0008-5472.CAN-11-3295

©2012 American Association for Cancer Research.

### Introduction

Percent mammographic breast density (PMD) is defined as the proportion of a mammographic image occupied by radiodense tissue (largely stromal and epithelial tissues, appearing as white regions on the mammogram) as opposed to nondense, fatty tissue (the darker regions of the image). PMD is one of the strongest known risk factors for breast cancer. Women with dense tissue in more than 75% of the breast have been shown to be at a 4- to 5-fold increased risk of breast cancer when compared with women who have mostly fatty breasts (1). The pathways through which the strong association between mammographic density and breast cancer is mediated are still unclear. Breast cancer has an estimated familial relative risk to first degree relatives of approximately 2 (2), and the heritability estimates for PMD range between 63% and 67% (3), hence it is plausible that these 2 traits share a common genetic basis.

A recent meta-analysis of 5 genome-wide association studies (GWAS) identified variants within intron 4 of the *ZNF365* gene that are associated with PMD, adjusted for age and body mass index (BMI; ref. 4). One of the *ZNF365* variants, rs10995190, had previously been shown in a GWAS to be associated with breast cancer in the same direction as the corresponding PMD

association (5). The meta-analysis also examined associations between 22 known breast cancer susceptibility loci and percent density. Two breast cancer single-nucleotide polymorphisms (SNP), rs2046210 (ESR1,  $P = 0.005$ ) and rs3817198 (LSP1,  $P = 0.04$ ), were found to be associated with percent density in the same direction as determined by the corresponding breast cancer associations, that is, the same allele was associated with higher density and with higher breast cancer risk (4). These findings suggest the existence of common genetic factors affecting density and breast cancer. Few, if any, other SNPs have been robustly confirmed as being associated with breast density (6).

Several recent studies have shown the potential value of using a score based on a combination of thousands of individual SNPs to predict disease status, and hence provide insights into the underlying genetic model of inheritance for that disease (for example, see refs. 7, 8). To further examine evidence for shared genetic basis between density and breast cancer, we identified a set of SNPs that showed the strongest evidence of association in the combined GWAS of breast density and examined their association with breast cancer risk in a separate GWAS, to determine whether the SNPs in combination were associated with breast cancer risk.

## Methods

### Breast cancer case-control study population

The present analysis was based on genotype data from the UK2 GWAS, as previously described in the work of Turnbull and colleagues (5). The cases were recruited from 23 clinical genetics centres in the United Kingdom through the Familial Breast Cancer Study (FBCS) and from oncology clinics in the United Kingdom through the Prospective study of Outcomes in Sporadic versus Hereditary (POSH) breast cancer study. Cases were preferentially selected to have at least 2 affected first- or second-degree relatives, and had been screened to exclude *BRCA1* or *BRCA2* mutations. The controls used in this study were genotyped as part of the Wellcome Trust Case Control Consortium (WTCCC) study and were drawn from the 1958 Birth Cohort (a population-based study in the United Kingdom of individuals born in 1 week in 1958) and from the U.K. National Blood Service (9). After quality control exclusions, the analysis was based on data from 3,628 cases and 5,190 controls. The constituent studies were all approved by their appropriate ethics committees.

### Breast cancer case-control genotyping

Genotypes for cases were generated by a custom Illumina Infinium 670k array, and controls were genotyped using an Illumina Infinium 1.2M array at the Wellcome Trust Sanger Institute, Cambridge, United Kingdom. Genotypes were called using the Illuminus algorithm; we used genotypes with a posterior probability more than 0.95. Quality controls steps are as described in the work of Turnbull and colleagues (5). Briefly, the analyses were based on individuals with a call rate of more than 97%. We eliminated cryptic duplicates and first-degree relative pairs based on identity-by-state (IBS) proba-

bilities, and individuals with non-European ancestry based on comparison with genotypes in HapMap phase II. We also excluded SNPs with a call rate less than 95% if their minor allele frequency (MAF) was more than 5%, SNPs with a call rate of less than 99% if the MAF was 5% or less, SNPs with MAF less than 1%, and SNPs whose frequencies departed from HWE at  $P < 0.00001$  in controls or  $P < 10^{-12}$  in cases. Cluster plots were inspected manually for the top SNPs as part of the UK2 GWAS. After exclusions, data were available for 531,158 SNPs that were successfully genotyped in 3,628 breast cancer cases and 5,190 controls (5).

The SNP data were used to impute genotypes for approximately 2.6 million SNPs in HapMap II CEU (release 22; build 36) with the program MACH (10). Chromosome X was imputed using HapMap II CEU, release 21 (build 35). Only SNPs for which the imputation  $r^2$  value was greater than 0.3 were included in the analyses, where  $r^2$  is the estimated squared correlation between imputed and true genotypes.

### Mammographic density measurements and genotyping

Associations between SNPs and density were based on the meta-analysis of 5 GWAS (totaling 4,887 individuals) previously reported by the MODE consortium (4). The 5 studies varied in their proportions of premenopausal women (0%–46%) and in their proportions of women using hormone replacement therapy at the time of their mammogram (9%–46%). All studies estimated density with the CUMULUS program (11). Percent density (dense area expressed as a proportion of total breast area), adjusted for age, BMI, and population stratification (using principal components analysis), was taken as the phenotype. All women were of self-described European descent and the majority (89%) were postmenopausal at the time of mammogram. Each study used multivariate adjustments on square root or natural logarithm transformed percent density before using appropriate regression methods to generate estimates of per-allele effects. Imputation was then used to generate statistics for approximately 2.6M SNPs, based on HapMap II CEU as a reference. One of the studies selected samples based on the upper and lower extremes of the density distribution, treated density as a binary trait and estimated per-allele ORs using logistic regression, whereas the remaining studies used samples unselected for density and analyzed density as a continuous trait using linear regression. Because the effect size estimates for the different studies were not directly comparable, the meta-analysis was conducted by deriving signed  $Z$  scores for each SNP for each study and deriving a test statistic by summing the  $Z$  scores weighted by the square root of the sample size (4).

Written informed consent was given by all participants. The individual studies were approved by The Committee on the Use of Human Subjects in Research at Brigham and Women's Hospital (Boston, MA); The Norfolk and Norwich Hospital Ethics Committee (Norwich, Norfolk, UK); the Ethical Review Board at Karolinska Institutet (Stockholm, Sweden); the Mayo Clinic Institutional Review Board (Rochester, MN); and the University Health Network (Toronto, ON, Canada), respectively.

### The Sisters in Breast Screening study

The Sisters in Breast Screening (SIBS) study of mammographic breast density was not included in the MODE meta-analysis of 5 GWAS (although it was included as a replication set for the most significant SNP in that study; ref. 4), and so it was possible to test the association of the polygenic risk score (PRS) with percent breast density using this set of women. The SIBS study is described in the work of Kataoka and colleagues (12). Briefly, families were identified through the National Health Service breast screening program in the United Kingdom. Eligibility was restricted to families in which 2 or more female blood relatives (sisters, half sisters, first cousins, or aunt–niece) had had mammographic screening. The study was approved by the local research ethical committee. Study recruitment commenced in October 2002. The current analysis was limited to families whose data including mammographic density measurements were completed by July 2007.

For each participant, all available mammograms were retrieved from the local screening unit and were digitized. The mammograms were scanned using the Array 2905 Laser Film Digitizer and the program DICOM ScanPro Plus Version 1.3E (Array Corporation), with 50- $\mu$ m pixel resolution and 12-bit digitization, and an absorbance of 4.7. For each individual, we aimed to collect the earliest and most recently available mammograms. Mammographic density was measured with the CUMULUS program (11). Mammograms were analyzed in a random order, and the reader was blinded to the sequence of the mammograms and to the visual density evaluation that had been done by another reader.

This analysis was based on 1,145 women from 563 families who had been genotyped as part of an ongoing GWAS. Samples were genotyped by Illumina using the Illumina HumanCytosNP-12 platform. Of 1,160 genotyped women, 6 were excluded on the basis of non-European ethnicity, and the twin with the lower call rate was excluded for each of the 9 pairs of monozygotic twins. The quality control steps were as for the UK2 GWAS; after exclusions genotypes were available for 184,838 SNPs. Genotype imputation was carried out in the same way as for the UK2 GWAS.

### Statistical methods

All analyses were done within R and used the GenABEL package (13) implemented in R. A list of the most significant 10% of imputed SNPs (236,090 SNPs) from the density meta-analysis was intersected with the complete list of SNPs genotyped in the UK2 GWAS (531,158 SNPs). The 50,899 SNPs found on both lists were used in the analyses. The same exercise was carried out for other cutoff points of density SNPs between 1% and 10%. As a sensitivity analysis, we repeated the analyses of the 1% to 5% cutoff points using the intersection of the imputed density SNPs with the imputed SNPs in the UK2 GWAS.

To derive an overall test statistic, we assumed that each SNP associated with density would confer an approximately proportional effect on breast cancer risk (as would be expected if density were an intermediate phenotype). Specifically, we assumed that the per-allele log-relative risk of breast cancer conferred by each SNP would be (approximately) proportional to the per-allele mean density difference. The aim was to create

a more informative score than would be obtained by simply summing the number of unweighted risk alleles. We further assumed that the combined effect of SNPs was additive for density and log additive for breast cancer risk (as has been observed for susceptibility SNPs identified to date).

Under this model, the log-odds of breast cancer for individual  $i$  would be related to the density SNPs using a PRS of the form:

$$\text{PRS}_i = \sum_{k=1}^N \beta_k G_{ik}$$

where  $\beta_k$  is the per-allele change in mean density for SNP  $k$  and  $G_{ik}$  is the genotype for SNP  $k$  in individual  $i$  (0,1, or 2) in the breast cancer GWAS. However, because the meta-analysis was done by combining  $P$  values, the effect sizes were only available as signed  $Z$  scores ( $Z_k$ ) rather than parameter estimates ( $\beta_k$ ). Although the above test statistic could be based on the  $Z$  scores rather than the  $\beta_k$ , this would be inefficient because for a given  $P$  value, a rarer SNP (lower MAF) would have had a higher effect estimate compared with a more common SNP (higher MAF), and hence a higher predicted relative risk of breast cancer.

For a standard linear regression analysis of density against SNP  $k$ , the standard error of  $\beta_k$  is given approximately by:

$$\text{SE}(\beta_k) \sim \sigma / \sqrt{2nq_k(1-q_k)}$$

where  $n$  is the sample size,  $q_k$  and  $1-q_k$  are the allele frequencies, and  $\sigma$  is the (residual) trait variance in the density study. For simplicity, we approximated the allele frequencies  $q_k$  from the density studies by the allele frequencies  $p_k$  in the breast cancer study. Hence:

$$Z_k = \beta_k / \text{SE}(\beta_k) \sim \beta_k \sqrt{2np_k(1-p_k)} / \sigma$$

so that for a given sample size, as here,  $\beta_k$  is proportional to  $Z_k / \sqrt{p_k(1-p_k)}$ . (The ranked density SNPs were based on a meta-analysis of imputations carried out by each group to the same reference panel, and so we would expect  $n$  to be the same for all SNPs, except in the case of SNPs that were poorly imputed in one or more studies.)

The PRS that we used for each individual was, therefore, of the form:

$$\text{PRS}_i = \sum_{k=1}^N Z_k G_{ik} / \sqrt{p_k(1-p_k)}$$

In computing the score, missing genotypes were replaced with the value 0, equivalent to assuming that the individual has a common homozygote genotype at that SNP (individuals with  $\geq 3\%$  of genotypes missing had been excluded).

In practice, only a subset of the SNPs will be associated with density (the remainder being false positives) and the effect sizes will often be overestimated. Nevertheless, provided that some of the associated SNPs are also associated with risk, the composite score,  $\text{PRS}_s$ , should be associated with breast cancer risk. To test this hypothesis, PRS was included as the exposure variable in a logistic regression model with



breast cancer case-control status as the outcome. This analysis was repeated for different cutoff points of the top density SNPs (1%–10%). For the sensitivity analysis using imputed SNPs in the breast cancer GWAS (rather than just the observed SNPs) the genotypes  $G_{ik}$  were replaced by the estimated allele doses.

To evaluate the probability of obtaining a significant association at any cutoff point by chance, a permutation  $P$  value was obtained by randomly permuting the case-control status 5,000 times and repeating the association tests with the PRS calculated for each cutoff point. The number of permutations for which the  $P$  value was less than or equal to the smallest actual  $P$  value were counted to estimate an empirical  $P$  value. The use of a permutation approach should make the test robust to the nonindependence of SNPs within the same linkage disequilibrium block.

To quantify the predictive value of the risk score, we calculated ORs by percentiles of PRS, using logistic regression.

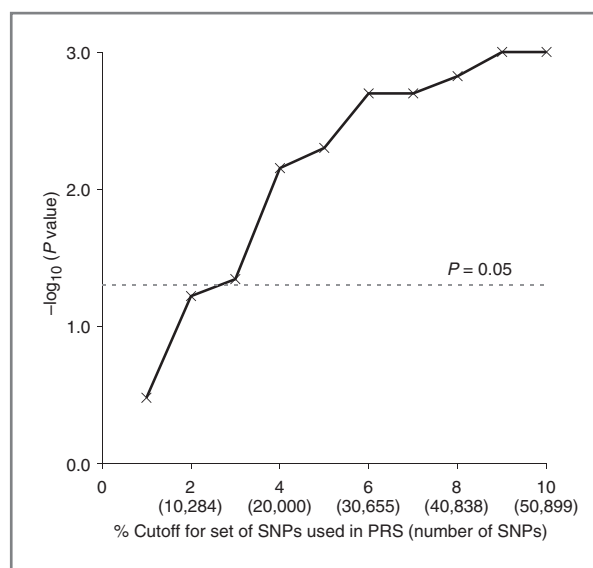
To assess the association of the PRS with breast density, we intersected the SNPs used to calculate the PRS in the case-control study (50,899 SNPs) with the list of SNPs successfully imputed in the SIBS study. Thus, the PRS for the SIBS women were calculated using 49,912 SNPs. Percent densities for up to 4 mammograms per woman (both breasts and 2 time points) were adjusted for age at mammogram and BMI, and the mean over the 4 values (or over all available images, if less than 4) was used. Linear regression was carried out to assess the association between the PRS and adjusted percent density, allowing for the nonindependence of relatives.

Statistical analyses were conducted using R and Stata version 10.0. All  $P$  values are 2 sided.

## Results

This study examined the association between breast cancer risk in 3,628 cases and 5,190 controls and a PRS based on a woman's genotypes at the top approximately 5,000 to 50,000 ranked SNPs from a genome-wide meta-analysis of mammographic breast density. The PRS based on the top 10% of density SNPs ranged between  $-5,274$  and  $6,982$  (mean =  $355$ , SD =  $1,594$ ). The mean 10% PRS was  $421$  (SD =  $1,582$ ) among the cases and  $310$  (SD =  $1,600$ ) in the controls ( $P = 0.0012$ ; Student  $t$  test). A one-point increase in the PRS was associated with an increase in percent breast density of  $6.9 \times 10^{-4}$  percentage points ( $P = 0.014$ ) in the SIBS study.

The significance of the association between PRS and breast cancer risk for different cutoff points of significant density SNPs are shown in Fig. 1. The PRS based on the top 1% of density SNPs (5,222 SNPs) was not significantly associated with breast cancer ( $P = 0.33$ ) but the association was significant at the 5% level for all cutoff points from 3% to 10%, being most significant for cutoff points between 6% and 10% ( $P = 0.002$  for the 6% cutoff point;  $P = 0.001$  for the 10% cutoff point; Fig. 1). For example, an increase of 1,000 PRS points (for the 10% cutoff point) was associated with a 4.5% increase in breast cancer risk [95% confidence interval (CI), 1.7%–7.3%]. Permutation analysis done across all 10 cutoff points (5,000 permutations) gave an overall empirical  $P$  value of 0.003.



**Figure 1.** Significance of the mammographic breast density PRS for the prediction of breast cancer risk, according to percent cutoff point of SNPs used in the PRS. Significance levels were obtained using unconditional logistic regression.

The analysis was also repeated using the imputed genotype dosages for the top 1% to 5% of density SNPs. The results followed a similar pattern as seen with genotyped SNPs. The score based on the top 5% of imputed SNPs was significantly associated with breast cancer risk at  $P = 0.01$  (Supplementary Fig. S1). Having established that the use of imputed SNPs did not change the broad pattern of the results, we did not extend the imputed analysis beyond the top 5% of SNPs for computational reasons.

We next categorized individuals by their risk score based on the top 10% of genotyped density SNPs (50,899 SNPs). We chose to use the scores based on the 10% cutoff point for convenience, although we note that the 10% scores are only slightly better than those obtained using the top 6% of SNPs. Women in the lowest decile of the PRS distribution were at a significantly lower risk of breast cancer than women with higher scores, with women in the top decile of PRS having an estimated OR = 1.31 (95% CI, 1.08–1.58) compared with the bottom decile (Table 1). There does not appear to be a steady gradient in risk, with women between the 10th and 25th percentiles, the 25th and 75th percentiles, and the 75th and 95th percentiles of the PRS distribution all having an approximately 19% higher risk than the lowest 10% of women. The group of SIBS women with the highest adjusted breast density were actually those between the 75th and 90th percentiles of PRS distribution ( $P = 0.014$  versus the lowest decile), but the SDs are large, given the relatively small numbers (Table 1).

## Discussion

In this study, we tested the hypothesis that common genetic factors contribute to the association between mammographic density and breast cancer. We calculated a PRS based on the signed  $Z$  scores for the set of SNPs showing the strongest

**Table 1.** Breast cancer ORs estimated for individuals within different percentiles of the PRS calculated using the top 10% (50,899) of SNPs associated with PMD

Percentiles of the PRS	UK2 study				SIBS study	
	No. of cases <sup>a</sup>	No. of controls <sup>a</sup>	OR (95% CI)	<i>P</i>	No. of women <sup>b</sup>	Mean PMD <sup>c</sup> (SD)
<10%	306	519	<i>Baseline comparison group</i>		115	-1.78 (11.8)
10%–25%	547	779	1.19 (1.00–1.42)	0.055	172	-1.42 (13.0)
25%–75%	1,830	2,595	1.20 (1.03–1.39)	0.022	572	-0.094 (12.7)
75%–90%	544	778	1.19 (0.99–1.42)	0.062	172	2.01 (13.8)
>90%	401	519	1.31 (1.08–1.59)	0.0058	114	0.95 (15.2)

<sup>a</sup>Percentiles taken from the PRS distribution in UK2 controls.

<sup>b</sup>Percentiles taken from the distribution of the PRS in the SIBS study (values were not exactly comparable with those in the UK2 study because of the slightly different set of SNPs for which data were available).

<sup>c</sup>Mean PMD, adjusted for age at mammogram and BMI (SD).

association with density in a meta-analysis of 5 breast density GWAS and examined whether this risk score was predictive of breast cancer status. We found that a risk score based on the top 10% of SNPs was significantly associated with breast cancer risk, whereas the risk score based on the top 1% was not. Consistent with this result women whose SNP profile is such as to give them a risk score in the top decile of the PRS distribution had a 31% increased risk of breast cancer compared with women in the lowest decile ( $P = 0.0058$ ), although the effect on breast cancer risk appears only to be detectable at either extreme of the polygenic allele distribution.

The aim was not to identify the specific SNPs that are associated with both traits; we note that we observed only 2 SNPs that were associated with both PMD and breast cancer with  $P$  values less than  $10^{-4}$ , one of which has been previously described (rs10995190; refs. 4, 5) and the other SNP (rs10509168) is also in the *ZNF365* gene (the 2 are not strongly correlated,  $r^2 = 0.13$ ). Our results suggest that more such SNPs should be reliably identifiable through larger meta-analyses of GWAS. Exclusion of SNPs rs10995190 and rs10509168 in *ZNF365* and all correlated SNPs ( $r^2 > 0.2$ ) made no material difference to the results (data not shown).

This study is the first to provide direct evidence from genotyping that mammographic breast density has a polygenic mode of inheritance, in line with other quantitative human traits, such as height and BMI (14). Using a Mendelian Randomization approach (15) with the PRS as the instrumental variable, we predicted that a 1,000-point increase in the PRS would be associated with an approximately 1.4% increase in breast cancer risk (95% CI, 0.041%–4.7%), based on our observed effect of the PRS on breast density and assuming a 1% increase in density is associated with a 2% increase in breast cancer risk (16, 17). This prediction is slightly smaller than the observed OR of 1.045 per 1,000-point increase in PRS (95% CI, 1.017–1.073;  $P = 0.001$ ). This supports the hypothesis that percent breast density is causally related to breast cancer (if the observed effect had instead been nonsignificant we would have concluded that the frequently observed association between density and breast cancer was a consequence of residual confounding). As the observed effect was slightly

larger than that predicted it might be the case that, in addition to their effect on density, these SNPs act pleiotropically to also increase breast cancer risk via other pathways, or that CUMULUS-estimated PMD does not fully capture the most relevant aspects of breast composition, and thus the PRS underestimates the true effects of these SNPs. The MODE meta-analysis included density results from 2 breast cancer case-control studies (hence, 1,324 of 4,877 women in MODE were patients with breast cancer), which may have resulted in some confounding by breast cancer status, which in turn could have inflated the observed effect of the density PRS on breast cancer risk. We were not able to reanalyze the meta-GWAS adjusting for case-control status, but the original report noted that for the most significant SNP (rs10995190) neither adjusting for case-control status nor excluding the cases changed the significance, suggesting that confounding was not present, at least for this SNP (4). We should also bear in mind that the set of cases used in this study were preferentially selected for an early age at breast cancer diagnosis and/or a positive family history of breast cancer, both of which may have upwardly biased the magnitude of the observed association (5).

One might have anticipated that the PRS based on the top 1% of density SNPs would show the most significant association with breast cancer risk. In fact, however, the significance of the association increased as further SNPs were added to the score, until about the 6% cutoff point, beyond which there was little improvement (a pattern broadly similar to that seen in the Multiple Sclerosis polygenic score study ref. 18). This suggests the existence of large numbers of SNPs associated with both traits, and that these are not limited to those variants showing the strongest associations with density. The flattening out of the graph beyond the 6% cutoff point may indicate a lack of power to detect variants with smaller shared effects, rather than an absence of such variants. Although one of the strengths of this study is the large number of samples in the density meta-GWAS from which the SNP rankings were obtained, the lower significance seen for the top 1% of SNPs may nevertheless point to a lack of power in the density study, resulting in imprecision in the ranking; an ongoing meta-analysis incorporating additional individual density studies may improve the precision

with which SNPs are ranked. It should also be noted that this study only considered SNPs associated with PMD. It is possible that the absolute volume of dense tissue in the breast is more relevant for breast cancer risk, and hence that a PRS based on SNPs associated with absolute dense area (as a hypothesized better surrogate for volume) would show stronger evidence of a shared genetic basis with breast cancer risk.

To the best of our knowledge, this is the first such study to use SNPs ranked on the basis of their association with a quantitative trait to predict risk of a related disease. Other groups have used similar approaches to obtain disease-risk information from SNPs that are not individually significant, with mixed results. For example, using WTCCC data, Evans and colleagues (8) found that disease-specific polygenic scores gave only limited discrimination between cases and controls of the same disease for bipolar disorder, coronary heart disease, hypertension, Crohn's disease, rheumatoid arthritis, and type II diabetes and that the discriminatory ability of the type I diabetes score was markedly reduced after the exclusion of SNPs in and around the MHC (a region known to be associated with type I diabetes). However, separate studies of multiple sclerosis and schizophrenia both found significant associations between disease risk and scores based on multiple SNPs, with the schizophrenia-derived polygenic score also predicting the risk of bipolar disease, but not of the 6 nonpsychiatric diseases in the WTCCC study described earlier (7, 18). Our results are in contrast to those of Machiela and colleagues (19) who, using a related approach, found no association between an unweighted PRS constructed using ranked SNPs from a breast cancer GWAS with breast cancer risk, after excluding 13 established breast cancer SNPs. Witte and Hoffmann also found that a similar polygenic model did not significantly predict breast cancer status (20). This presumably reflects the low power of both studies due to small training and testing sets (<1,200 cases and controls in total for each) and shows the need for large studies when considering variants with small effects. The power of our study was, in addition, boosted by the enrichment of the set of breast cancer cases for a positive family history of the disease.

Theoretically, the association we observed could result from population stratification, if there were strata with higher density and higher breast cancer risk. However, this is unlikely as the GWAS for density were conducted in several populations in which there was little evidence of population stratification. Moreover, the individual risk scores were computed using many, largely unlinked tag-SNPs that are highly unlikely to be related to population structure. In the present analysis, for individuals with missing genotype data, the missing genotypes were replaced with the value 0, which is equivalent to assuming that the individual has a common homozygote genotype at that SNP. An alternative approach would have been to substitute the missing genotype count with the mean count of risk alleles for the respective SNP. However, this should not make any material difference as only individuals with at least 97% call rates were included in the breast cancer GWAS and the call rates were not significantly different among cases and controls (5). A recent study reported no significant difference in their results with or without substitution of missing genotypes (21).

This analysis shows directly that a suitably chosen set of SNPs that do not reach genome-wide significance levels individually for one trait can nevertheless predict a genetically related trait when used in combination. This method could be used to look for shared genetic influences between other risk factors and breast cancer, or indeed between any pair of correlated heritable traits.

The recent identification of common SNPs associated with breast cancer risk indicate that susceptibility to breast cancer is at least partly polygenic (22). Our results suggest that a much larger number of common variants are involved, and, moreover that there is at least some overlap between the set of genetic variants involved in breast density and those involved in breast cancer risk. This observation in turn confirms that the well-established correlation between the 2 traits is genuine, and not the result of uncorrected confounding. It is not possible at this stage to determine whether the observed effect is attributable to variants which increase cancer risk via the creation of a dense breast, or to variants which act pleiotropically to affect both traits, or to a mixture of both types of variants. The identification of the relevant variants should help to elucidate the biologic mechanisms underlying breast density and the reasons for its association with breast cancer.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Acknowledgments

The authors thank Tina Audley for her role in coordinating the SIBS study. They also acknowledge the use of DNA from the British 1958 Birth Cohort collection.

#### Grant Support

MODE breast density GWAS: This study was supported by Public Health Service Grants CA131332, CA087969, CA049449 from the National Cancer Institute, NIH, Department of Health and Human Services. The NHS breast cancer cases and controls were genotyped with support from the National Cancer Institute's Cancer Markers of Susceptibility (CGEMS) initiative. Data evaluation of mammograms and analysis of the EPIC-Norfolk study was supported by Cancer Research (UK). The SASBAC study was supported by Märit & Hans Rausing's Initiative against Breast Cancer, NIH, Susan Komen Foundation and Agency for Science, Technology and Research of Singapore (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.

UK2 breast cancer GWAS: This work was supported by the Wellcome Trust and by Cancer Research UK. C. Turnbull is funded by a Medical Research Council Clinical Research Fellowship. The samples were collected and screened for *BRCA* mutations through funding from Cancer Research UK; US Military Acquisition (ACQ) Activity, Era of Hope Award (W81XWH-05-1-0204), and the Institute of Cancer Research (UK). This study makes use of data generated by the Wellcome Trust Case Control Consortium (WTCCC) 2. A full list of the investigators who contributed to the generation of the data is available from the WTCCC website. The British 1958 Birth Cohort collection was funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Funding for this project was provided by the Wellcome Trust under award 085475.

J.S. Varghese is funded by The Cambridge Commonwealth Trust and a Cambridge Overseas Research Scholarship. D.F. Easton is a Principal Research Fellow of Cancer Research (UK).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 7, 2011; revised January 5, 2012; accepted January 15, 2012; published OnlineFirst January 19, 2012.

## References

- 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.
- Collaborative Group on Hormonal Risk Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001;358:1389–99.
- 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.
- 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.
- 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.
- Kelemen LE, Sellers TA, Vachon CM. Can genes for mammographic density inform cancer aetiology? *Nat Rev Cancer* 2008;8:812–23.
- Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;460:748–52.
- Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 2009;18:3525–31.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–78.
- Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009;10:387–406.
- Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. *Phys Med Biol* 1994;39:1629–38.
- 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.
- Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;23:1294–6.
- Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011;43:519–25.
- Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008;5:e177.
- Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *J Natl Cancer Inst* 1995;87:670–5.
- Mitchell G, Antoniou AC, Warren R, Peock S, Brown J, Davies R, et al. Mammographic density and breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Cancer Res* 2006;66:1866–72.
- Bush WS, Sawcer SJ, de Jager PL, Oksenberg JR, McCauley JL, Pericak-Vance MA, et al. Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am J Hum Genet* 2010;86:621–5.
- Machiela MJ, Chen CY, Chen C, Chanock SJ, Hunter DJ, Kraft P. Evaluation of polygenic risk scores for predicting breast and prostate cancer risk. *Genet Epidemiol* 2011;35:506–14.
- Witte JS, Hoffmann TJ. Polygenic modeling of genome-wide association studies: an application to prostate and breast cancer. *OMICS* 2011;15:393–8.
- Li S, Zhao JH, Luan J, Ekelund U, Luben RN, Khaw KT, et al. Physical activity attenuates the genetic predisposition to obesity in 20,000 men and women from EPIC-Norfolk prospective population study. *PLoS Med* 2010;7:e1000332.
- Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 2002;31:33–6.



# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Mammographic Breast Density and Breast Cancer: Evidence of a Shared Genetic Basis

Jajini S. Varghese, Deborah J. Thompson, Kyriaki Michailidou, et al.

*Cancer Res* 2012;72:1478-1484. Published OnlineFirst January 19, 2012.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/0008-5472.CAN-11-3295](https://doi.org/10.1158/0008-5472.CAN-11-3295)

**Supplementary Material** Access the most recent supplemental material at:  
<http://cancerres.aacrjournals.org/content/suppl/2012/01/19/0008-5472.CAN-11-3295.DC1>

**Cited articles** This article cites 22 articles, 3 of which you can access for free at:  
<http://cancerres.aacrjournals.org/content/72/6/1478.full#ref-list-1>

**Citing articles** This article has been cited by 5 HighWire-hosted articles. Access the articles at:  
<http://cancerres.aacrjournals.org/content/72/6/1478.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cancerres.aacrjournals.org/content/72/6/1478>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.