

ARTICLE

Received 14 Mar 2014 | Accepted 17 Sep 2014 | Published 24 Oct 2014

DOI: 10.1038/ncomms6303

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. Paterson⁵, Jingmei Li⁶, Gretchen L. Gierach⁷, Christopher Scott⁸, Jennifer Stone⁹, Julie A. Douglas¹⁰, Isabel dos-Santos-Silva¹¹, Pablo Fernandez-Navarro^{12,13}, Jajini Verghese^{3,4,14}, Paula Smith^{3,4}, Judith Brown^{3,4}, Robert Luben³, Nicholas J. Wareham¹⁵, Ruth J.F. Loos^{15,16}, John A. Heit¹⁷, V. Shane Pankratz⁸, Aaron Norman⁸, Ellen L. Goode⁸, Julie M. Cunningham¹⁸, Mariza deAndrade⁸, Robert A. Vierkant⁸, Kamila Czene⁶, Peter A. Fasching^{19,20}, Laura Baglietto^{21,22}, Melissa C. Southey²³, Graham G. Giles^{21,22}, Kanaan P. Shah¹⁰, Heang-Ping Chan²⁴, Mark A. Helvie²⁴, Andrew H. Beck²⁵, Nicholas W. Knoblauch²⁵, Aditi Hazra^{1,2,26}, David J. Hunter^{1,2,26}, Peter Kraft^{1,2,27}, Marina Pollan^{12,13}, Jonine D. Figueroa⁷, Fergus J. Couch^{8,18}, John L. Hopper²², Per Hall⁶, Douglas F. Easton^{3,4,28}, Norman F. Boyd²⁹, Celine M. Vachon^{8,*} & Rulla M. Tamimi^{1,2,26,*}

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 ($P < 5 \times 10^{-8}$) 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 and *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.

¹Program in Genetic Epidemiology and Statistical Genetics, Harvard School Of Public Health, Boston, Massachusetts 02115, USA. ²Department of Epidemiology, Harvard School Of Public Health, Boston, Massachusetts 02115, USA. ³Centre for Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK. ⁴Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. ⁵Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada M5G 1X8. ⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden. ⁷Hormonal and Reproductive Epidemiology Branch, National Cancer Institute, Bethesda, Maryland 20850, USA. ⁸Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, USA. ⁹Centre for Genetic Origins of Health and Disease, University of Western Australia, Perth, Western Australia 6009, Australia. ¹⁰Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ¹¹Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. ¹²Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid 28029, Spain. ¹³Consortium for Biomedical Research in Epidemiology and Public Health (CIBER en Epidemiología y Salud Pública—CIBERESP), Madrid 28029, Spain. ¹⁴Plastic Surgery Unit, Royal Free Hospital, London NW3 2QG, UK. ¹⁵Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge CB1 8RN, UK. ¹⁶The Icahn School of Medicine at Mount Sinai, The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, New York, New York 10029, USA. ¹⁷Division of Cardiovascular Disease, Department of Medicine, Mayo Clinic, Rochester, Minnesota 55905, USA. ¹⁸Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota 55905, USA. ¹⁹Department of Gynecology and Obstetrics, Erlangen University Hospital, Friedrich Alexander University of Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, 910 54 Erlangen, Germany. ²⁰Division Hematology/Oncology, Department of Medicine, University of California at Los Angeles, David Geffen School of Medicine, Los Angeles, California 90024, USA. ²¹Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria 3004, Australia. ²²Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria 3010, Australia. ²³Department of Pathology, University of Melbourne, Melbourne, Victoria 3010, Australia. ²⁴Department of Radiology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ²⁵Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02115, USA. ²⁶Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ²⁷Department of Biostatistics, Harvard School Of Public Health, Boston, Massachusetts 02115, USA. ²⁸Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. ²⁹Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute, Toronto, Ontario, Canada M5G 2M9. * These authors jointly supervised this work. Correspondence and requests for materials should be addressed to R.M.T. (email: rulla.tamimi@channing.harvard.edu).

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 cancer¹, 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}; however, 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) genome-wide association studies (GWAS) of DA, NDA and PD, respectively. We identify genome-wide significant ($P < 5 \times 10^{-8}$) loci for DA (*AREG*, *ESR1*, *ZNF365*, *LSP1/TNNT3*, *IGF1*, *TMEM184B* and *SGSM3/MKL1*), NDA (8p11.23) and PD (*PRDM6*, 8p11.23 and *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 11 studies with GWAS data (Methods, Supplementary Note 1) comprising a total of 7,916 women. Study subjects were predominantly postmenopausal women of European ancestry participating in the Markers of Density consortium. Mammographic density was measured using CUMULUS⁹ (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 single-nucleotide polymorphism (SNP), we combined study-specific *P* values and direction of association using the METAL software¹⁰. We assessed 200 promising SNPs for replication in up to 10,379 women from 11 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 ($P < 5 \times 10^{-8}$) 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 that strongly correlate with our top SNP rs10034692 in this region have previously been associated with breast size¹¹. Although we observed the strongest association for rs10034692, another SNP (rs12642133) located 116 kb away and in weak linkage disequilibrium (LD) with rs10034692 ($r^2 = 0.16$, $D' = 1.00$) also reached genome-wide significance (Supplementary Table 4). We investigated these two SNPs further in 6,624 women from the NHS (Nurses' Health Study), BBCC (Bavarian Breast Cancer Cases and Controls), MCBCC (Mayo Clinic Breast Cancer Study) and MMHS (Mayo Mammography Health Study) studies for whom we had individual-level genotype data. Both SNPs were associated with DA in this data set when analysed separately ($\beta = -0.16$,

Table 1 | SNPs associated with mammographic DA, NDA and PD.

Lead SNP	Chr: position*	Gene [†]	Alleles [‡]	MAF [§]	Stage	Z-Score	P-value	Stage 1 + 2			
								β (s.e.)	Z-Score	P-value	P(Het) [¶]
<i>SNPs associated with mammographic DA</i>											
rs10034692	4:75638651	<i>AREG</i>	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	<i>ESR1</i>	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	<i>ZNF365</i>	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	<i>LSP1</i>	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	<i>IGF1</i>	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	<i>TMEM184B</i>	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	<i>SGSM3, MKL1</i>	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}				
<i>SNPs associated with mammographic NDA</i>											
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}				
<i>SNPs associated with mammographic PD</i>											
rs186749	5:122482204	<i>PRDM6</i>	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	<i>TMEM184B</i>	C/A	0.40	1	-4.75	1.99×10^{-6}	-0.10 (0.02)	-5.86	4.68×10^{-9}	0.82
					2	-3.54	0.0004				

DA, dense area; MAF, minor allele frequency; NDA, non-dense area; PD, percent density; SNP, single-nucleotide polymorphism.

*HG version 18.

[†]Includes nearby genes.

[‡]Major allele (reference allele)/minor allele (effect allele).

[§]MAF as in the 1000 Genomes project.

^{||} β and s.e. estimates were obtained using fixed effects meta-analysis of cross-sectional studies (that is, studies that analysed density phenotypes as a quantitative trait) in unrelated individuals.

Mammographic density phenotypes are square-root-transformed.

[¶]*P* value for heterogeneity between studies.

$P=0.0002$ for rs10034692 and $\beta=0.17$, $P=9 \times 10^{-6}$ 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 ($r^2=0.87$, $D'=1.00$) and in moderate LD with SNPs previously associated with breast size¹¹.

The rs10995190 SNP in the *ZNF365* region has been associated with both PD⁶ and breast cancer risk¹⁴; however, 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 ($r^2=0.08$, $D'=0.36$ with rs10995190) and rs10733779 ($r^2=0.11$, $D'=1.00$ with rs10995190) showed genome-wide significant associations with DA. After adjusting for rs10995190, the associations for both rs1949359 ($P=4.4 \times 10^{-5}$ before and $P=0.008$ after adjustment) and rs10733779 ($P=1.9 \times 10^{-6}$ before and $P=0.002$ after adjustment) were attenuated. Additional analyses in larger data sets will be necessary to determine whether there are multiple independent SNPs associated with DA 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 risk¹⁷ and is hypothesized to be involved in breast development. Indeed, circulating levels of IGF1 are associated with breast cancer risk¹⁸.

We also confirmed previous findings⁸ 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 ($r^2=0.24$, $D'=0.82$). Both these SNPs have been associated with breast cancer risk, and the recently published iCOGS¹⁹ 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 ($r^2=0.34$, $D'=0.71$) located in the *PLA2G6* gene has previously been associated with cutaneous nevi²⁰. 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/MKLI* region) is in moderate LD ($r^2=0.41$, $D'=0.76$) with rs6001930 that has been previously associated with breast cancer¹⁹. We also observed several nearby SNPs located in the *TNRC6B* and *MKLI* 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 $P=2.4 \times 10^{-23}$), 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* (ref. 6) and 12q24 (ref. 7) loci (Supplementary Figs 10 and 11). Through replication analysis (Supplementary

Table 6), we identified three new loci ($P<5 \times 10^{-8}$) 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 the regulation of endothelial cell proliferation, survival and differentiation. Interestingly, we observed a borderline association ($P=2.6 \times 10^{-7}$) 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 rs10995190. As with DA, analysis in larger data sets will be needed to assess the possibility of multiple independent SNPs in this region.

We used data from the ENCODE²⁴ 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 ($r^2 \geq 0.8$) 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/MKLI* 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 site²⁵. 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. On the basis of 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/MKLI*) was associated with *SGSM3* expression in both adipose tissue and LCL ($P<0.0001$). We also examined whether any of these SNPs (or proxies) were associated with transcript levels in breast cancer tumours using data from The Cancer Genome Atlas²⁷ (TCGA). We conducted both *cis* (within 1 Mb of the transcription start or end site) and *trans* (genome-wide) expression quantitative trait loci (eQTL) analyses. 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 whether 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/MKLI*)^{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

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*	Z-score, mammographic density association (density phenotype [†])	Breast cancer association OR (95% CI)	P [‡] breast cancer association
Dense area	rs10034692	4	<i>AREG</i>	A/G	-6.35 (DA)	0.99 (0.97-1.01)	0.31
	rs12665607	6	<i>ESR1</i>	T/A	5.64 (DA)	1.20 (1.16-1.23)	1.48×10^{-30}
	rs3817198	11	<i>LSP1</i>	T/C	6.47 (DA)	1.07 (1.05-1.09)	2.09×10^{-13}
	rs703556	12	<i>IGF1</i>	A/G	-6.27 (DA)	0.94 (0.90-0.99)	0.02
	rs17001868	22	<i>SGSM3/MKL1</i>	A/C	-7.33 (DA)	1.10 (1.08-1.13)	1.19×10^{-15}
Percent density	rs186749	5	<i>PRDM6</i>	G/A	5.96 (PD)	1.02 (1.01-1.04)	0.009
Dense area and percent density	rs10995190	10	<i>ZNF365</i>	G/A	-8.26 (DA)	0.86 (0.84-0.88)	1.50×10^{-37}
	rs7289126	22	<i>TMEM184B</i>	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}

CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism.

*Reference allele/effect allele.

[†]DA, dense area; PD, percent density; NDA, non-dense area.

[‡]P value based on a fixed effects meta-analysis. Estimates and s.e.'s for each breast cancer study were calculated using logistic regression.

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; however, 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 cm² for DA, 8.58 cm² for NDA and 0.77% for PD. 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: DA, NDA and PD. 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 DA and breast cancer risk but in opposite directions. Interestingly, there is accumulating data that *MKL1* may have both tumour-inhibiting and tumour-promoting roles depending on the cellular context. Recently, it was shown that the *MKL1* signalling pathway was activated in estrogen receptor (ER) – cell lines and silenced in ER + cell lines²⁹. In addition, the *MKL1* breast cancer

SNP has been shown to be associated with triple-negative breast cancer³⁰. rs17001868 has also been associated with *SGSM3* (involved in the 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 tissue²⁹ 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 DA 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 DA 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 (for example, age and menopausal status). Similarly, rs7816345 was also associated with apparent opposing directions on NDA and breast cancer risk. In addition, 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 cancer³¹, while postmenopausal BMI is positively associated with breast cancer³².

There are some weaknesses with our study that should be mentioned. First, we used the HapMap project as imputation panel that 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 Genomes³³ 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 whether 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 pathways³⁴. 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 (Singapore and Sweden Breast Cancer Study (SASBAC)), Mayo Clinic (MBCFS (Minnesota Breast Cancer Family Study), MAYO VTE, MCOCS, MMHS, MCBCS), University Health Network, Toronto, Canada (TOR), Eastern Multicentre Research Ethics Committee (Sisters in Breast Screening (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 Committee (Melbourne Collaborative Cohort Study (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 12 GWAS of mammographic density (Supplementary Note 1). For DA and NDA, we had GWAS data from 11 studies, and for PD we had GWAS data from 12 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 iCOGS¹⁹ 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 ~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 MAF < 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 and 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 matrix³⁷. 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 DA (AMDTSS) were defined as 'cases' and women in the bottom 10% percentile were defined as 'controls'. As the included data from the NHS were generated using two different genotyping platforms, they were analysed as two separate studies. Similarly, data from the SASBAC were obtained through two separate genotyping efforts and therefore were analysed 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 software¹⁰. 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 an 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 genotyping¹⁹ 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 ($r^2 \geq 0.80$) 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 MMHS, the MCBCS and the MCCS using a customized iSelect array. We excluded subjects with call rates $< 95\%$ ($n = 44$) and two subjects (out of 204 included duplicates) who 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 was 5,647 women.

Assessment of regulatory functions for identified SNPs. We used the ENCODE²⁴ data to assess whether any of the identified mammographic density SNPs or their proxies ($r^2 \geq 0.8$ in 1000 Genomes CEU population) are located in regulatory regions. Look-ups were made using the HaploReg³⁸ and RegulomeDB³⁹ softwares. We also investigated whether identified mammographic density SNPs or their proxies were associated with gene expression *in cis* in adipose tissue and LCLs in the MuTHER⁴⁰ data by accessing the GeneVar²⁶ database. To further explore the regulatory properties of the mammographic density SNPs, we conducted eQTL analyses on mammographic density SNPs and their proxies ($r^2 \geq 0.8$) using data from 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 oestrogen receptor-positive cases and 19,105 transcripts among 177 oestrogen 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 $N \times \log(N)$ 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 iCOGS¹⁹ + GAME-ON breast cancer GWAS meta-analysis. The GAME-ON meta-analysis^{13,19,43,44} can be found at (<http://gameon.dfc.harvard.edu>) and is based on 11 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.

References

- Vachon, C. M. *et al.* Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res.* **9**, 217 (2007).
- Pettersson, A. *et al.* Nondense mammographic area and risk of breast cancer. *Breast Cancer Res.* **13**, R100 (2011).
- Pettersson, A. *et al.* Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *J. Natl Cancer Inst.* **106**, 5 (2014).
- Boyd, N. F. *et al.* Heritability of mammographic density, a risk factor for breast cancer. *New Engl. J. Med.* **347**, 886–894 (2002).
- Stone, J. *et al.* The heritability of mammographically dense and nondense breast tissue. *Cancer Epidemiol. Biomark. Prev.* **15**, 612–617 (2006).

6. Lindstrom, S. *et al.* Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. *Nat. Genet.* **43**, 185–187 (2011).
7. Stevens, K. N. *et al.* Identification of a novel percent mammographic density locus at 12q24. *Hum. Mol. Genet.* **21**, 3299–3305 (2012).
8. Vachon, C. M. *et al.* Common breast cancer susceptibility variants in LSP1 and RAD51L1 are associated with mammographic density measures that predict breast cancer risk. *Cancer Epidemiol. Biomark. Prev.* **21**, 1156–1166 (2012).
9. Byng, J. W. *et al.* Symmetry of projection in the quantitative analysis of mammographic images. *Eur. J. Cancer Prev.* **5**, 319–327 (1996).
10. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
11. Eriksson, N. *et al.* Genetic variants associated with breast size also influence breast cancer risk. *BMC Med. Genet.* **13**, 53 (2012).
12. Long, J. *et al.* Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet.* **8**, e1002532 (2012).
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. *Hum. Mol. Genet.* **21**, 5373–5384 (2012).
14. Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42**, 504–507 (2010).
15. Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **41**, 324–328 (2009).
16. Cai, Q. *et al.* Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. *Hum. Mol. Genet.* **20**, 4991–4999 (2011).
17. Hunter, D. J. *et al.* A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat. Rev. Cancer* **5**, 977–985 (2005).
18. Key, T. J., Appleby, P. N., Reeves, G. K. & Roddam, A. W. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. *Lancet Oncol.* **11**, 530–542 (2010).
19. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 361e1–362ee (2013).
20. Nan, H. *et al.* Genome-wide association study identifies nidogen 1 (NID1) as a susceptibility locus to cutaneous nevi and melanoma risk. *Hum. Mol. Genet.* **20**, 2673–2679 (2011).
21. Zhang, M. *et al.* Association between Cutaneous Nevi and Breast Cancer in the Nurses' Health Study: A Prospective Cohort Study. *PLoS Med.* **11**, e1001659 (2014).
22. Kvaskoff, M. *et al.* Association between Melanocytic Nevi and Risk of Breast Diseases: The French E3N Prospective Cohort. *PLoS Med.* **11**, e1001660 (2014).
23. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361 (2013).
24. Consortium, E. P. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
25. Stacey, S. N. *et al.* Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. *PLoS Genet.* **6**, e1001029 (2010).
26. Yang, T. P. *et al.* Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**, 2474–2476 (2010).
27. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
28. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
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.* **390**, 34–44 (2014).
30. Purrington, K. S. *et al.* Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis* **35**, 1012–1019 (2014).
31. Baer, H. J., Tworoger, S. S., Hankinson, S. E. & Willett, W. C. Body fatness at young ages and risk of breast cancer throughout life. *Am. J. Epidemiol.* **171**, 1183–1194 (2010).
32. van den Brandt, P. A. *et al.* Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am. J. Epidemiol.* **152**, 514–527 (2000).
33. Genomes Project, CAbecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
34. Varghese, J. S. *et al.* Mammographic breast density and breast cancer: evidence of a shared genetic basis. *Cancer Res.* **72**, 1478–1484 (2012).
35. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62**, 1198–1211 (1998).
36. Amos, C. I. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54**, 535–543 (1994).
37. Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
38. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
39. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
40. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
41. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
42. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
43. Ghousaini, M. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat. Genet.* **44**, 312–318 (2012).
44. Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.* **45**, 398e1–398e2 (2013).

Acknowledgements

This study was supported by CA131332, CA087969, CA049449, CA128931, CA116201, CA075016, CA122340, CA122844, CA15083, CA089393, K22LM011931 and X01 HG005954 from NIH; genotyping services for the OOA study were provided by the Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096; The Breast Cancer Research Foundation, Breast Cancer Research Fund; Cancer Research UK; Märit & Hans Rausing's Initiative against Breast Cancer; Susan Komen Foundation; Agency for Science, Technology and Research of Singapore (A*STAR); David F. and Margaret T. Grohne Family Foundation; Campbell Family Institute for Breast Cancer Research; David F. and Margaret T. Grohne Family Foundation; Ontario Ministry of Health and Long Term Care; Fashion Footwear Charitable Foundation of New York/QVC Presents Shoes on Sale; FIS PI060386 from the Spain's Health Research Fund and EPY 1306/06 Collaboration Agreement between Astra-Zeneca and the Carlos III Institute of Health; Elizabeth C. Crosby Research Award, Gladys E. Davis Endowed Fund and the Office for Vice President of Research at the University of Michigan. EPIC-Norfolk was funded by research programme grant funding from Cancer Research UK and the Medical Research Council with additional support from the Stroke Association, British Heart Foundation, Department of Health, Research into Ageing and Academy of Medical Sciences. The SIBS study was supported by programme grant C1287/A10118 and project grants from Cancer Research UK (grant numbers C1287/8459). SEARCH is funded by a programme grant from Cancer Research UK (C490/A10124). The Polish Breast Cancer Study was supported (in part) by the Intramural Research Program of the National Institutes of Health, National Cancer Institute. The breast cancer meta-analysis is supported by the GAME-ON DRIVE (CA148065) and BCAC initiatives. We thank the BCAC, GAME-ON and DRIVE initiatives for generously sharing breast cancer association results for selected SNPs. We also thank the investigators in BCAC, PRACTICAL, CIMBA and OCAC for access to the iCOGS data for the replication analysis. BBCC was funded in part by the ELAN Programme of the Medical Faculty, Friedrich-Alexander University Erlangen-Nuremberg. We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. The OOA study investigators thank the members of the Amish community for their generous support and participation, the staff at the Amish Research Clinic for their dedicated recruitment and fieldwork efforts, the members of Dr Margarita Shultz's radiology clinic for their expert mammography services, the staff at the Center for Inherited Disease Research for their exceptional genotyping services, Drs Alan Shuldiner and Braxton Mitchell at the University of Maryland for their guidance and help with our fieldwork, and Terry Gliedt, Jennifer Greene Nidetz, Kristen Maas, Cris Van Hout, James MacDonald, Chris Plotts, Lubomir Hadjiiski and Chuan Zhou at the University of Michigan for their technical assistance with data management, entry and analysis and film digitization and scoring. The PBCS would like to thank Pei Chao and Michael Stagner from Information Management Services (Silver Spring, MD) for data management support; Laurie Burdette, Amy Hutchinson and Jeff Yuenger from the NCI Core Genotyping facility for genotyping support; the participants, physicians, pathologists, nurses and interviewers from participating centres in Poland for their efforts during fieldwork; Drs Jola Lissowska and Ewa Wesolowska from the Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland for their assistance with mammogram retrieval; Dr Norman Boyd from the University of Toronto for providing the mammographic density assessments; and Drs. Louise Brinton, Montserrat Garcia-Closas, Beata Peplonska and Mark Sherman for their contributions to the study design. We thank Paul Pharoah and the SEARCH and EPIC teams. The iCOGS project would not have been possible without the contributions of the following: Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog and Ken Offitt (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and

the staff of Mayo Clinic Genotyping Core Facility. The iCOGS infrastructure was funded by the following: the European Community's Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175; COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation and the Ovarian Cancer Research Fund. The results published here are in part based upon data generated by The Cancer Genome Atlas project established by the NCI and NHGRI (dbGaP Study Accession: phs000178.v8.p7). Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <http://cancergenome.nih.gov/>.

Author contributions

S.L., D.J.T., A.D.P., J.L., G.L.G., J.S., J.L.H., P.H., D.F.E., N.F.B., C.M.V. and R.M.T. designed the study. S.L., D.J.T., A.D.P., J.L., J.S., C.S., J.A.D., P.F.-N., J.V., P.S., A.H.B.

and N.W.K. performed the statistical analysis. J.A.D., I.D.-S.-S., J.B., R.L., N.J.W., R.J.F.L., J.A.H., V.S.P., A.N., E.L.G., J.M.C., M.D., R.A.V., K.C., P.A.F., L.B., M.C.S., G.G.G., K.P.S., H.-P.C., M.A.H., A.H., D.J.H., P.K., M.P., J.D.F., F.J.C., J.L.H., P.H., D.F.E., N.F.B., C.M.V. and R.M.T. provided samples and data. S.L., D.J.T., A.D.P., J.A.D., J.L.H., N.F.B., C.M.V. and R.M.T. drafted the manuscript. All authors contributed to the final paper.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Lindström, S. *et al.* Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk. *Nat. Commun.* 5:5303 doi: 10.1038/ncomms6303 (2014).

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

Sara Lindström, Deborah J. Thompson, Andrew D. Paterson, Jingmei Li, Gretchen L. Gierach, Christopher Scott, Jennifer Stone, Julie A. Douglas, Isabel dos-Santos-Silva, Pablo Fernandez-Navarro, Jajini Verghase, Paula Smith, Judith Brown, Robert Luben, Nicholas J. Wareham, Ruth J.F. Loos, John A. Heit, V. Shane Pankratz, Aaron Norman, Ellen L. Goode, Julie M. Cunningham, Mariza deAndrade, Robert A. Vierkant, Kamila Czene, Peter A. Fasching, Laura Baglietto, Melissa C. Southey, Graham G. Giles, Kaanan P. Shah, Heang-Ping Chan, Mark A. Helvie, Andrew H. Beck, Nicholas W. Knoblauch, Aditi Hazra, David J. Hunter, Peter Kraft, Marina Pollan, Jonine D. Figueroa, Fergus J. Couch, John L. Hopper, Per Hall, Douglas F. Easton, Norman F. Boyd, Celine M. Vachon & Rulla M. Tamimi

Nature Communications 5:5303 doi: 10.1038/ncomms6303 (2014); Published 24 Oct 2014; Updated 9 Sep 2015

The financial support for this Article was not fully acknowledged. The Acknowledgements should have included the following:

This study was supported by CA173785 from the National Institutes of Health.