Genome wide association study identifies a novel putative mammographic density locus at 1q12-q21

Pablo Fernandez-Navarro1,2, Anna González-Neira3, Guillermo Pita3, Ramón Díaz-Uriarte4, Leticia Tais Moreno3, María Ederra2,5, Carmen Pedraz-Pingarrón6, Carmen Sánchez-Contador7, José Antonio Vázquez-Carrete8, Pilar Moreo9, Carmen Vidal10, Dolores Salas-Trejo11, Jennifer Stone12, Melissa C. Southey13,14, John L. Hopper15,16, Beatriz Pérez-Gómez13,18, Javier Benitez13,17,18 and Marina Pollan1,2

1 Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain
2 Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública—CIBERESP), Spain
3 Human Genotyping Unit-CeGen, Spanish National Cancer Research Centre, Madrid, Spain
4 Department of Biochemistry, Universidad Autónoma de Madrid, Instituto de Investigaciones Biomédicas “Alberto Sols” (UAM-CSIC), Madrid, Spain
5 Navarre Breast Cancer Screening Programme, Public Health Institute, Pamplona, Spain
6 Castilla-Leon Breast Cancer Screening Programme, D.G. Salud Pública ID e I, Castilla y León, Spain
7 Balearic Islands Breast Cancer Screening Programme, Provincial Health Service, Balearic Islands, Spain
8 Galicia Breast Cancer Screening Programme, Regional Authority of Health, Galicia Regional Government, A Coruña, Spain
9 Aragon Breast Cancer Screening Programme, Health Service of Aragon, Zaragoza, Spain
10 Cancer Prevention and Control Unit, Catalan Institute of Oncology (ICO), Barcelona, Spain
11 Valencia Breast Cancer Screening Programme, Directorate Public Health & Centre for Public Health Research (CSISP), Valencia, Spain
12 Centre for Genetic Origins of Health and Disease (GOHaD), Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, Perth, Western Australia, Australia
13 Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia
14 Melbourne Collaborative Cohort Study (MC3S)
15 School of Population Health, Centre for Molecular Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Victoria, Australia
16 Department of Health Science and Services, School of Public Health, Seoul National University, Seoul, Korea
17 Human Genetics Group, Spanish National Cancer Research Centre (Centro Nacional de Investigaciones Oncológicas—CNIO), Madrid, Spain
18 Consortium for Biomedical Research in Rare Diseases (Centro de Investigación Biomédica en Red de Enfermedades Raras—CIBERER), Spanish National Cancer Research Centre, Madrid, Spain

Mammographic density (MD) is an intermediate phenotype for breast cancer. Previous studies have identified genetic variants associated with MD; however, much of the genetic contribution to MD is unexplained. We conducted a two-stage genome-wide association analysis among the participants in the “Determinants of Density in Mammographies in Spain” study, together with a replication analysis in women from the Australian MD Twins and Sisters Study. Our discovery set covered a total of 3,351 Caucasian women aged 45 to 68 years, recruited from Spanish breast cancer screening centres. MD was blindly assessed by a single reader using Boyd’s scale. A two-stage approach was employed, including a feature selection phase exploring 575,374 SNPs in 239 pairs of women with extreme phenotypes and a verification stage for the 183 selected SNPs in the remaining sample (2,873 women). Replication was conducted in 1,786 women aged 40 to 70 years old recruited via the Australian Twin Registry, where MD were measured using Cumulus-3.0, assessing 14 SNPs with a p value <0.10 in stage 2.

Key words: DDM-Spain, GWAs, mammographic density, MTMR11, OTUD7B
Abbreviations: BC: breast cancer; BMI: body mass index; DDM-Spain: Determinants of Density in Mammographies in Spain Study; LD: linkage disequilibrium; MD: mammographic density; OR: odds ratio; PC: principal component analysis; PHV1: peak in height growth velocities in infancy; PHV2: peak in height growth velocities in the puberty; SNAP: SNP Annotation and Proxy Search; SNP: single nucleotide polymorphisms

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Spain’s Health Research Fund (Fondo de Investigación Sanitaria); Grant numbers: FIS PI060386 and PIS09/01006; Grant sponsor: Collaboration Agreement between Astra-Zeneca and the Carlos III Institute of Health (Instituto de Salud Carlos III); Grant number: EPY 1306/06; Grant sponsor: Spanish Federation of Breast Cancer Patients; Grant number: FECMA 485 EPY 1170-10; Grant sponsor: Spanish MINECO; Grant number: BIO2009-12458
DOI: 10.1002/ijc.29299

History: Received 7 Apr 2014; Accepted 16 Oct 2014; Online 29 Oct 2014
Correspondence to: Pablo Fernandez-Navarro, Cancer and Environmental Epidemiology Unit, National Centre for Epidemiology, Carlos III Institute of Health, Avda. Monforte de Lemos 5 28029, Madrid, Spain, Tel.: +34-918222644, Fax: +34-913877815, E-mail: pfernandezn@isciii.es
Mammographic density (MD) reflects variations in the amounts of fat, stromal and epithelial tissue in the breast gland. The measurement of MD has been proposed as an intermediate phenotype for breast cancer susceptibility.\textsuperscript{1,2} Twin and family studies have shown that genetic variation accounts for at least 60% of MD variability.\textsuperscript{1,3,4}

The relation between breast cancer (BC) genetic variants and MD has been studied by different groups finding significant associations with several inter and intragenic SNPs.\textsuperscript{5–10} A pooled cross-sectional analysis\textsuperscript{11} of common breast cancer susceptibility variants in 14 independent loci found variants in \textit{LSP1} (rs3817198) and \textit{RAD51L1} (rs10483813) genes associated with mammographic density measures. Recently, two previous genome-wide association studies (GWAS)\textsuperscript{12,13} have identified new genetic variants associated with MD, one in gene \textit{ZNF365} (rs10995190) and other between genes TBX5 and TBX3 (rs1265507). However, these single nucleotide polymorphisms (SNPs) only account for a 0.5 to 1.3% of the MD variance, leaving much of the genetic contribution to MD unexplained.

In this study, we have done a two-stage GWAS\textsuperscript{14,15} inside a multicentre study of 3,574 Spanish women attending BC screening (DDM-Spain “Determinants of Density in Mammographies in Spain”), together with a replication analysis of our best hit in 1,786 women from the Australian Twins and Sisters Study of Mammographic Density.\textsuperscript{8} Furthermore, we analysed the genetic region in linkage disequilibrium (LD) around our best hit.

**Material and Methods**

**Two-stage GWAS (DDM-Spain): Subjects and mammograms**

DDM-Spain is a cross-sectional study, which aimed to identify genetic, reproductive and lifestyle characteristics associated with mammographic patterns/densities that might enhance the risk of developing breast cancer. The study was conducted from September 2006 to June 2007 and included women aged 45 years old and over who attended the regional Breast Cancer Screening Programmes. The characteristics of this study have been previously published.\textsuperscript{16–18} Exclusion criteria included evidence of previous breast or ovarian cancer, inability to answer the questionnaire, physical impairment to perform the mammogram and previous breast implants.

The study was approved by the Bioethics Committee of the Carlos III Institute of Health (Instituto de Salud Carlos III) (Madrid) and all subjects gave their written consent. The sample consisted of 3,574 women (range 497 to 536 per centre). Of the total sample, only Caucasian women (3,351) were genotyped and included in this study.

Menopausal status was self-reported and defined as absence of menstruation in the preceding 12 months and MD was measured from the craniocaudal mammogram of the left breast, using a visual scale that rates density in six categories (Boyd’s semi-quantitative scale), namely: A (0%); B (<10%); C (10–25%); D (25–50%); E (50–75%); and F (>75%). Mammographic density was assessed with high reproducibility, by a single, experienced radiologist on a blind, anonymous basis. For quality control purposes, a random sample of 374 mammograms was analyzed in duplicate showing a high concordance between both readings (weighted kappa value of 0.92).\textsuperscript{19}

**Two-stage GWAS (DDM-Spain): DNA extraction, genotyping**

Genomic DNA was extracted from saliva using Oragene DNA Collection Kit (DNAgenotek), and DNA was quantified using PicoGreen (Invitrogen Corp., Carlsbad, CA).

For the first stage, 239 pairs of women with extreme phenotypes (low/high) were selected: women with MD greater than 50% were individually matched with women with MD <10% who came from the same center, had a similar age (±2 years), menopausal status and BMI (±2 kg/m\(^2\)). In this first stage, DNA samples were genotyped with the Illumina Human610-Quad BeadChip platform (Illumina, San Diego, CA). After implementing a quality control to data where we checked the possible sample stratification, SNP genotype missing, SNP monomorphic status, SNP minor allele frequency and SNP “ilumina score” (see Supporting Information.

Finally, two genetic variants in high linkage disequilibrium with our best hit were studied using the whole Spanish sample. Evidence of association with MD was found for variant rs11205277 (OR = 0.74; 95% CI = 0.67–0.81; \(p = 1.33 \times 10^{-10}\)). In replication analysis, only a marginal association between this SNP and absolute dense area was found. There were also evidence of association between MD and SNPs in high linkage disequilibrium with rs11205277, rs11205303 in gene \textit{MTMR11} (OR = 0.73; 95% CI = 0.66–0.80; \(p = 2.64 \times 10^{-11}\)) and rs67807996 in gene \textit{OTUD7B} (OR = 0.72; 95% CI = 0.66–0.80; \(p = 2.03 \times 10^{-11}\)). Our findings provide additional evidence on common genetic variations that may contribute to MD.

**What’s new?**

Mammographic density, the ratio of breast tissue to fat in the breast, serves as a marker for breast cancer risk. Genes influence breast density, but it’s not yet known exactly how. In this paper, the authors searched for genetic loci that might contribute to mammographic density. Using a two-stage genome-wide study, they found three genetic variants that associated with breast density. This study is one of the largest yet conducted on the genetic component of breast density.
In the second stage, we genotyped 172 SNPs from the 183 SNPs selected in stage 1 (those with an Illumina Score >0.60, see Supporting Information File 1) in the remaining 2,873 Caucasian women with DNA available. Genotyping was performed, using VeraCode GoldenGate Genotyping Assay according to the protocols issued by Illumina (Illumina, San Diego, CA). Data were analysed, using GenomesStudio software for genotyping clustering and calling. Interplate and intraplate replicate samples were genotyped. In addition, genotyping data from CEPH trios (Coriell Cell Repository, Camden, NJ) were also included across the plates to identify Mendelian inconsistencies. Only 156 SNPs of the 172 SNPs genotyped in this second stage were analyzed after implementing a quality control (see Supporting Information File 1).

To ensure results observed in VeraCode GoldenGate Genotyping Assay were consistent and to confirm there were no technical problems, some of the genotyped variants were analyzed in a subset of random samples across plates using alternative genotyping techniques (Infinium or Taqman assays) and no discordant results were observed.

Replication analysis (Australian Twins Study): Subjects and mammograms
The replication data consisted of 1,786 women aged 40 to 70 years old who participated in the Australian Mammographic Density Twins and Sisters Study between 2004 and 2009 recruited via the Australian Twin Registry.

Mammographic density measurements (absolute dense area and percentage of dense area) were performed using Cumulus 3.0, a computer-assisted thresholding technique, by three independent operators, and a telephone-administered questionnaire captured self-reported sociodemographic information (i.e., weight, height, smoking history and cessation of menstruation) (see details in Ref. [8]).

Replication analysis (Australian MD Twins and Sisters Study): DNA extraction, genotyping
DNA was extracted from blood samples and TaqMan assays (Applied Biosystem) with fluorescent allele-specific probed were used to genotype the genetic variants (see details).

Statistical analysis. Two-stage GWAS (DDM-Spain)
In the first stage (“feature selection stage”) which used only extreme MD phenotypes, two statistical procedures were used to select our best candidates. First, for each SNP we fitted a logistic regression model adjusted by menopausal status (yes/no), age and body mass index (BMI) and assumed a log-additive genetic model. Second, we used a Bayesian-inspired penalised maximum likelihood approach, to examine the additive contribution of SNPs to MD risk. This method selects a subset of SNPs that best predict disease status, while controlling the type I error of selected SNPs and attempts to identify multiple causal variants, instead of testing one SNP at a time. Briefly, this method uses a logistic regression where each SNP is a covariate, and the problem is one of selecting the best covariates (the best SNPs). This variable selection is carried out using a penalized maximum likelihood approach, with Bayesian stochastic search, where the coefficient for each SNP has a sharp prior at 0 (i.e., no effect). The final output from the method is a subset of SNPs that best predict disease status.

To select the SNPs for the stage 2, SNPs had to meet the following criteria: a p value ≤0.001 in the first analysis and a non-cero posterior estimate (with a prior precision of 30—see “Bayesian approach” in Supporting Information File 1). A total of 183 SNPs fulfilled these criteria.

The second stage used the remained 2,873 women with MD classified, as mentioned before, in six categories. Due to the small number of women in the extreme categories, MD was reclassified into four categories: 0–10%, 10–25%, 25–50% and >50%. To test the association between MD and the selected SNPs, we performed an ordinal logistic regression adjusting for recruitment center, menopausal stratus (yes/no), age at mammography (continuous) and Body Mass Index (BMI) (restricted quadratic splines), assuming a log-additive genetic model. We checked the proportionality assumption in our ordinal logistic regression models using the Brant test.

Population stratification was assessed in both stages by performing a Principal Component (PC) analysis and adjusting our results for the first two components (PCs) from this analysis (Supporting Information Files 2 and 3). The inclusion of these PCs did not affect the initial results, suggesting that population stratification was minimal, if present at all (data not shown). The quantile-quantile and Manhattan plots for the feature selection stage (stage 1) are shown in Supporting Information Files 4 and 5. The genetic inflation factor for the stage 1 was 1.020.

These statistical analyses were performed using R software; analyses were carried out in duplicate independently at the National Epidemiology Center and the Spanish National Cancer Research Center to avoid any error in execution.

Statistical analysis. Replication analysis (Australian MD Twins and Sisters Study)
The associations between MD and the 14 SNPs which showed a p value <0.10 in the second stage in the Spanish analysis were estimated in the Australian sample using linear regression, assuming an additive genetic model and adjusting for age and BMI. The residual variance was assumed to be constant, and the covariance between sisters was allowed to differ according to whether they were monozygotic twins, dizygotic twins, or nontwin sisters. Parameters and confidence intervals were estimated by maximum likelihood
We performed an association analysis between MD and the genetic variants in LD ($r^2 > 0.5$) with our best hit to determine whether other SNPs in the region also corroborated the association with MD in this region and to identify possible casual variants located in nearby genes. For this purpose, we used the “SNP Annotation and Proxy Search” (SNAP) to identify these genetic variants. The inputs used in the proxy search function of SNAP are the following: “1,000 Genomes Pilot 1” as SNP data set, CEU population as “panel” chosen, and a distance of 500 kilobases between the query SNP and the proxy SNP. Two genetic variants identified using SNAP and not included in the Illumina Human610-Quad BeadChip platform were genotyped in the whole sample of the DDM-Spain study (3,351 women) using Taqman assays (Applied Biosystem).

The association between these SNPs and MD in the whole sample was assessed using the same statistical approach already described (ordinal logistic regression adjusting for recruitment center, menopausal status, age and BMI).

### Results

#### Two-stage GWAS (DDM-Spain)

| Variable | Stage 1 | Stage 2 |
|----------|---------|---------|
| N        | 478     | 2,873   |
| Mean age at mammography (yr) | 57.31 (4.82) | 55.88 (5.55) |
| Mean body mass index (kg/m²) | 26.10 (2.21) | 28.35 (5.25) |
| No. post-menopausal women (%) | 416 (87) | 2,155 (75) |
| Mammographic density categories | | |
| 1 (<10%) | 239 (50%) | 585 (20%) |
| 2 (10–25%) | Nap | 676 (24%) |
| 3 (25–50%) | Nap | 1,070 (37%) |
| 4 (>50%) | 239 (50%) | 542 (19%) |

Data shown as n(%) or mean (SD). Abbreviations: Nap = not applicable.

The association analysis in stage 2 are shown in Table 2 for those SNPs with p values lower than 0.05. Only one SNP, rs11205277, in chromosome 1, displayed a significant association with MD after Bonferroni correction ($p = 5.57 \times 10^{-5}$), showing an odds ratio (OR) (per minor allele increase and likelihood of higher density category) of 0.66–0.85; CI = 0.69–0.85). To test the consistency of this association across different characteristics of our women that are in turn associated with MD, Figure 1 shows the association analysis between MD and the genetic variant rs11205277 stratified by categories of age at mammography, BMI, menopausal status and recruiting center. No visual evidence of a heterogeneous effect across different strata of these variables was found.

### Association analysis of Linkage disequilibrium region

We used the “SNP Annotation and Proxy Search” (SNAP) to identify these genetic variants. The inputs used in the proxy search function of SNAP are the following: “1,000 Genomes Pilot 1” as SNP data set, CEU population as “panel” chosen, and a distance of 500 kilobases between the query SNP and the proxy SNP. Two genetic variants identified using SNAP and not included in the Illumina Human610-Quad BeadChip platform were genotyped in the whole sample of the DDM-Spain study (3,351 women) using Taqman assays (Applied Biosystem).

The association between these SNPs and MD in the whole sample was assessed using the same statistical approach already described (ordinal logistic regression adjusting for recruitment center, menopausal status, age and BMI).

### Replication analysis (Australian MD Twins and Sisters Study)

The majority of the Australian MD Twins and Sisters Study participants (69%) were postmenopausal, with a mean age of 55 years old and a mean BMI of 26. None of the 14 SNPs analyzed in the Australian MD Twins and Sisters Study showed a statistically significant association with the absolute dense area or percentage dense area. However marginal associations between dense area and our best hit, rs11205277 and the genetic variants rs12607966 and rs7230021 were found (estimates = −0.13, 0.14 and 0.13 and p values = 0.07, 0.08 and 0.08, respectively) (see Table 3).

### Association analysis of linkage disequilibrium region

Regional LD plot in Figure 2 shows that rs11205303 and rs67807996 are in LD ($r^2 = 0.87$ and 0.58, respectively) with rs11205277. Rs11205303 is a missense variant (M159V) located in the exon 6 of the MTMR11 gene (Myotubularin related protein 11) and rs67807996 is located at 12,579 bps upstream of the start codon of the OUTD7B gene (Zinc finger protein Cezanne).

Both genetic variants, rs11205303 and rs67807996 were significant associated with MD ($p = 2.64 \times 10^{-11}$ and $2.03 \times 10^{-11}$, respectively) (OR (per minor allele increase and likelihood of higher density category) = 0.73; CI = 0.66–0.80 and OR (per minor allele increase and likelihood of higher density category) = 0.72; CI = 0.66–0.80, respectively) using stage 1 and stage 2 samples combined (3,351 women) (see Table 4). The association between rs11205277 and MD (stage 1 and 2 samples combined) was also significant ($p$ value = $1.33 \times 10^{-10}$; OR (per minor allele increase and
The genotype frequencies for these three SNPs by mammographic density category are shown in Table S3 (Supporting Information File 8).

The stratified analyses of the SNPs rs11205277, rs11205303 and rs67807996 in the whole sample (3351 women) across categories of age, BMI, menopausal status and recruiting center are shown in Figure S5 of the Supporting Information File 9. Again, there was no evidence of any interaction between these variables and the SNPs.

**Discussion**

Using a two-stage GWAs strategy together with a replication analysis we described a novel significant association between a susceptibility genetic variant (rs11205277) at 1q12-q21 and MD adjusting for age, BMI and menopausal status. We also have shown an association between MD and two SNPs, in high LD with our best hit, one of them (rs11205303) located in the MTMR11 gene and the other (rs67807996) located in the cis-regulatory region of the OTUD7B gene.

The most consistent result in our study was a decrease in the risk of MD associated with the variant rs11205277, located in the MTMR11 gene.
Table 3. Results of the association analysis in the Replication analysis (Australian MD Twins and Sisters Study)

| Polymorphism locus | Chromosome | Position | MAF | HWE | Alleles | Genotype | Outcome |
|-------------------|------------|----------|-----|------|---------|----------|---------|
| rs12650052        | 4          | 87,357,762 | 15  | 0.84 | C/T     | 1293     | N = 1,786 |
| rs11205277        | 1          | 148,159,496 | 42  | 0.52 | A/G     | 600      | p = -0.13 |
| rs6771728         | 2          | 55,345,558  | 0   | 0.93 | A/G     | 1,779    | p = -0.15 |
| rs4666755         | 10         | 96,622,243  | 43  | 0.8  | C/T     | 592      | p = -0.03 |
| rs4944707         | 11         | 86,831,745  | 48  | 0.11 | A/G     | 470      | p = -0.11 |
| rs1402704         | 11         | 75,617,248  | 30  | 0.23 | C/T     | 865      | p = -0.00 |
| rs1622354         | 2          | 2,935,003   | 29  | 0.44 | G/T     | 905      | p = 0.00  |
| rs1571526         | 13         | 97,716,489  | 35  | 0.98 | T/C     | 766      | p = 0.04  |
| rs12607966        | 18         | 49,921,278  | 27  | 0.83 | C/T     | 943      | p = 0.14  |
| rs17164879        | 5          | 122,891,860 | 21  | 0.81 | C/T     | 1,125    | p = -0.14 |
| rs1738822         | 6          | 47,203,703  | 21  | 0.03 | G/A     | 1,112    | p = 0.05  |
| rs7230021         | 18         | 49,792,489  | 28  | 0.13 | G/A     | 929      | p = 0.13  |
| rs9861551         | 3          | 60,664,516  | 4   | 0.93 | G/A     | 1,639    | p = -0.07 |
| rs1780330         | 1          | 162,969,178 | 50  | 0    | C/T     | 2        | p = 1.33  |

1 Position = position of SNP (36.3 NCBI reference genome build).
2 MAF = Minor allele frequency (%).
3 HWE = p value of Hardy-Weinberg equilibrium test.
4 Alleles = major allele/minor allele.
5 Wild = number of participants with homozygous wild type.
6 Heter = number of participants with heterozygous variant.
7 Homo = number of participants with homozygous variant.
8 N = total number of participants genotyped.
9 Estimate = estimated coefficient for the genetic variant in a linear regression, assuming an additive genetic model and adjusting for age and BMI.
10 SE = coefficient standard error.
11 p = p value for the linear regression.
an intergenic position at chromosome 1. No prior genome-wide association study has shown an association between this SNP and MD.\textsuperscript{12,13} The function of this genetic region is unclear. However, Gudbjartsson \textit{et al.}\textsuperscript{25} found a high correlation of this SNP with height in Caucasian people, showing that the Histone class 2A, \textit{MTMR11}, \textit{SV2A}, and \textit{SF3B4} genes are neighbouring the loci. In fact, BMI, a measurement closely related with height, is also associated with MD.\textsuperscript{26,27} Another study\textsuperscript{28} found that rs11205277 was associated with two stages of height growth, the peak height velocities in infancy (PHV1) and the puberty (PHV2). Moreover Lei \textit{et al.}\textsuperscript{29} confirmed the association of this SNP with stature. Finally, Okada \textit{et al.}\textsuperscript{30} and Zhao \textit{et al.}\textsuperscript{31} did not find any association between this SNP and height measurements in Japanese subjects and European American children, respectively, suggesting that there are different genetic backgrounds determining height across populations. In our study, we also found a positive association between height and rs11205277 ($\beta$ estimate = 0.42 cm, $p = 0.004$). Moreover, there is an association between height (cm) and mammographic density (OR (per height (cm) increase and likelihood of higher density category) = 1.05; CI = 1.03–1.06; $p = 7.29 \times 10^{-17}$). Finally, we also found that the association between rs11205277 and mammographic density (OR (per minor allele increase and likelihood of higher density category) = 0.74; CI = 0.67–0.81; $p$ value = 1.53 $\times 10^{-10}$) was still significant after additional adjustment for height.

The replication analysis partially confirmed our results, while no association between the percentage of MD and rs11205277 was found, the analysis using the absolute dense area showed a marginal association going in the same direction as ours. The lack of a strong association in the Australian study may be partly explained by a different genetic background of these two populations (Spaniards vs. Australians), but can also be related with the different method of MD assessment used in each study, differences in study designs or in genotyping platforms. Regarding MD assessment, in DDM-Spain, an experienced radiologist read the images using a semi-quantitative scale, while the Australian study used CUMULUS, a computer-assisted method that provides a quantitative measure of the percentage area as well as the absolute area occupied by dense tissue. While visual reading evaluates the image as a whole and may overestimate MD, semi-automatic methods compute the percentage of dense tissue taking the whole area limited by the skin into account, including the subcutaneous fat tissue that is not part of the mammary gland.\textsuperscript{26,27} This, in turn, may imply a greater agreement between visual assessment and the absolute dense area that is not influenced by the amount of subcutaneous fat.

Rs11205303 is in high LD with rs11205277 ($r^2 = 0.87$). This SNP is located in the gene “myotubularin related protein 11” (\textit{MTMR11}) at 1q12-q21, which has been mentioned previously in relation with height.\textsuperscript{25} This gene has a phosphatase activity\textsuperscript{12,33} and no prior GWAS study has found a significant association between this gene and MD. However, it is altered by amplification in the 7% of 482 human breast tumours cases analysed in the study of the Cancer Genome Atlas Network.\textsuperscript{34} Moreover, Lucci \textit{et al.}\textsuperscript{35} found \textit{MTMR11} gene altered by analyzing expression in breast cancer cells.

On the other hand, rs11205303 is also highly correlated with rs67807996 ($r^2 = 0.70$). This variant falls within the upstream regulatory region of the \textit{OTUD7B} gene, making likely that it could affect gene expression by altering gene

![Figure 2. Regional LD plot of rs11205277. Created by the web application of SNP Annotation and Proxy Search (SNAP).](image-url)
regulatory sequences. However, the data in track “Chromatin State Segmentation by HMM from ENCODE/Broad” of USCS GenomeBrowser (http://genome.ucsc.edu/) do not show any regulation category for this SNP, and we have not found any promoter, enhancer marks or evidence of variation in expression or protein binding in “Haploreg” database (http://www.broadinstitute.org/mammals/haploreg/haploreg.php). Moreover, we found a minimal evidence of binding a regulatory element examining ENCODE data available in RegulomeDB (http://regulomedb.org/index).

OTUD7B gene encodes the Zinc finger protein Cezanne, a known deubiquitination enzyme that inhibits NF-jB activity and contributes to cancer progression by deubiquinination of EGFR. This is the first time this gene has been related with MD; however, OTUD7B gene is altered by amplification in the 7.9% of 482 human breast tumours cases analysed in the study of the Cancer Genome Atlas Network. In fact the role of the genetic variation in the ZNF365 gene, other zinc finger protein (365), in MD variability and also in cancer susceptibility has been previously described.

The relation between genetic variants and MD has been studied by different groups finding significant associations with several inter and intragenic SNPs. In relation with the pooled cross-sectional analysis of common breast cancer susceptibility variants and MD, the SNP rs10483813 in gene RAD51L1 was not assessed in our study, while the association between the rs3817198 variant in gene LSP1 with MD was not statistically significant (p = 0.2676) although the OR was in the same direction (OR (per minor allele increase and likelihood high vs. low density) = 1.16; CI = 0.89–1.51) (see Supporting Information File 7). Finally, the two previous genome-wide association studies (GWAS) have identified new genetic variants associated with MD, one in gene ZNF365 (rs10995190) and another between genes TBX5 and TBX3 (rs1265507). In our study, although the ORs were in the same direction as those previously reported for the two SNPs, rs10995190 and rs1265507 (OR (per minor allele increase and likelihood high vs. low density) = 0.58; CI = 0.40–0.85 and OR (per minor allele increase and likelihood high vs. low density) = 0.96; CI = 0.74–1.24, respectively), only the first of them reached statistical significance (p = 4.45 × 10^-3) in stage 1 (see Supporting Information File 7).

The differences in the study design, samples (population-based vs. family-based sample), population, genotyping platforms and statistical approach between our analysis and the studies mentioned could be related with the lack of replication. The Winner’s Curse could also be playing an important role in the weak replication found. Most importantly, the main end-point, mammographic density, was visually assessed in the Spanish study using a semi-quantitative scale of six categories by an experienced reader. This classification has been associated with the risk of subsequent breast cancer in Spain. In other studies, density was measured using a computer-assisted tool.

Table 4. Results of the association analysis for the best hit and the two selected SNPs in the LD region including all women from DDM-Spain (3,351 women, stages 1 and 2 samples combined)

| Polymorphism | Locus | Chromosome | Position | Minor allele frequency (%) | Hardy-Weinberg equilibrium (HWE) | Alleles | OR (95% CI) | p-value |
|--------------|-------|------------|----------|---------------------------|-------------------------------|--------|-------------|---------|
| rs11205277   | 1     | 148,159,496| A/G      | 0.01                       | 1.33 × 10^-10                 | 0.74   | 0.67–0.81   | 0.14    |
| rs67807996   | 1     | 148,261,889| G/A      | 0.01                       | 3.38 × 10^-10                 | 0.72   | 0.66–0.80   | 0.10    |
| rs11205303   | 1     | 148,173,037| T/C      | 0.03                       | 3.28 × 10^-11                 | 0.66–0.80 | 0.10    |

Note: MAF = Minor allele frequency (%), HWE = p-value of Hardy-Weinberg equilibrium test, OR = odds ratio, ordinal logistic regression model adjusted for recruitment centre, age at mammography, body mass index (BMI) and menopausal status.

Epidemiology

Int. J. Cancer: 136, 2427–2436 (2015) © 2014 UICC
(Cumulus) in which the reader highlights the edge of the breast and moves the pointer to select what he/she considers is the dense zone and density is automatically computed. This implies that density values in both studies are not directly comparable, not only due to the semi-quantitative nature of the Spanish data, but most importantly due to the differences in percentage of density obtained by radiologists and Cumulus (see Ref. [27 for a discussion regarding the advantages and disadvantages of different methods of MD assessment). In fact, the research team involved in the DDM-Spain study has recently compared MD estimates obtained with two computer-assisted tools (Cumulus and DM-Scan) with those provided by the radiologist in digital mammograms. Our results confirmed that both computer-assisted methods yielded lower values of density, something in agreement with previous articles describing an underestimation of percent of density using semi-automatic methods. While visual reading may overestimate density due to the evaluation of the image as a whole and the impossibility to take into account non-dense pixels included in a dense area, computer tools include the subcutaneous fat tissue as part of the breast, which implies an overestimation of the non-dense area. In summary, it is difficult to join the information available in the studies and variations in density should be assessed within each study.

Our work is one of the largest epidemiological studies conducted to date analyzing genetic determinants of mammographic density and including pre and post-menopausal women. Previous GWAs were based on sample sizes ranging from 1,66213 to 4,877.12 DDM-Spain study recruited Spanish women attending population-based Breast Cancer screening programs. Our participants seem to be representative of the entire Spanish female population of the same age range.31 Finally, the two-stage design using extreme phenotypes in the first stage has shown to be a cost-efficient strategy for identification of new genetic variants.52

The measurement of density was performed visually by a single radiologist using categorical scales and might be considered a limitation. While the use of quantitative methods has been recommended,27 they are not immune from subjectivity, have been only validated for analog mammograms and underestimate the percentage of density in digital mammograms.7,27 In our study, three of the seven screening centers used full-field digital images. Furthermore, the use of a single radiologist, with a good intra-observed reproducibility,19 minimized the measurement error in our study. Moreover, using the same procedure, we have found a significant association between visual estimation of MD and subsequent cancer risk.44

In summary, polymorphisms rs11205277, rs11205303 in gene MTMR11, and rs67807996 in gene OTUD7B were associated with MD under a log-additive model providing additional evidence that common genetic variations contribute to MD. The mechanisms whereby these genetic regions may affect breast density and, ultimately, cancer development are unclear. However, these variants are amplified in some breast cancer cases and one of them seems to be related with height, suggesting that they may influence dense tissue across growth and eventually modulate breast cancer risk.

Acknowledgements

The authors thank the participants in the DDM-Spain study for their contribution to breast cancer research. SNP genotyping services were provided by the Spanish “Centro Nacional de Genotipado” (CEGEN-ISCIID) (“www.cegen.org”). Other members of the Human Genotyping Unit-CeGen (Spanish National Cancer Research Centre) participating in the study were Belen Herráez y Rosario Alonso. Other members of DDM-Spain: Virginia Lope, Anna Cabanes, Nuria Aragonés and Gonzalo López-Abente are with the National Center for Epidemiology, Instituto de Salud Carlos III, Madrid, Spain; Virginia Lope, Anna Cabanes, Beatriz Pérez-Gómez, Nuria Aragonés, Gonzalo López-Abente, Nieves Ascunce, Juana Vidán, and Jesus Vioque are with Consortium for Biomedical Research in Epidemiology & Public Health (CIBER en Epidemiología y Salud Pública, CIBERESP), Spain; Nieves Ascunce and Juana Vidán are with Navarra Breast Cancer Screening Programme, Public Health Institute, Pamplona, Spain; Carmen Santamaria, Montserrat Coroju and Ana Belén Fernández are with Galicia Breast Cancer Screening Programme, Regional Authority of Health, Galicia Regional Government, Galicia, Spain; Maria Pilar Moreno and Soledad Abad are with Aragon Breast Cancer Screening Programme, Health Service of Aragon, Zaragoza, Spain; Francisco Ruiz-Perales, Maria Soledad Laso, Josefa Miranda-Garcia and Manuela Alcaraz are with Valencia Breast Cancer Screening Programme, General Directorate Public Health, Valencia, Spain and Centro Superior de Investigación en Salud Pública (CSISP), Valencia, Spain; and Jesus Vioque is also with Universidad Miguel Hernandez, Alicante, Spain; Francisco Casanova is with Castilla-Leon Breast Cancer Screening Programme, D.G. Salud Pública ID e I, Castilla y León, Spain; Francisca Collado-García is with Baleares Islands Breast Cancer Screening Programme, Health Promotion for women and childhood. General Directorate Public Health and Participation, Regional Authority of Health and Consumer Affairs, Baleares Islands, Spain. R.D.-U. work was partially supported by project from the Spanish MINECO (BIO2009–12458).

References

1. Boyd NF, Rommens JM, Vogt K, et al. Mammographic breast density as an intermediate phenotype for breast cancer. Lancet Oncol 2005;6:798–808.
2. Boyd NF, Martin LJ, Yaffe MJ, et al. Mammographic density and breast cancer risk: current understanding and future prospects. Breast Cancer Res 2011;13:223.
3. Stone J, Dite GS, Gunasekara A, et al. The heritability of mammographically dense and nondense breast tissue. Cancer Epidemiol Biomarkers Prev 2006;15:612–17.
4. Vachon CM, Sellers TA, Carlson EE, et al. Strong evidence of a genetic determinant for mammographic density, a major risk factor for breast cancer. Cancer Res 2007;67:8412–18.
5. Tamimi RM, Cox D, Kraft P, et al. Breast cancer susceptibility loci and mammographic density. Breast Cancer Res 2008;10:R66.
6. Wookcott CG, Maskarinec G, Haiman CA, et al. Association between breast cancer susceptibility loci and mammographic density: the Multiethnic Cohort. Breast Cancer Res 2009;11:R10.
7. Lee E, Haiman CA, Ma H, et al. The role of established breast cancer susceptibility loci in mammographic density in young women. Cancer Epidemiol Biomarkers Prev 2008;17:258–60.
8. Odeffy F, Stone J, Gurrin LC, et al. Common genetic variants associated with breast cancer and mammographic density measures that predict disease. Cancer Res 2010;70:1449–58.
9. Fernandez-Navarro P, Pita G, Santamaría C, et al. Association analysis between breast cancer and...
genetic variants and mammographic density in a large population-based study (Determinants of Density in Mammographies in Spain) identifies susceptibility loci in TOX3 gene. *Eur J Cancer* 2012;49:474–81.

10. Ellingford-Dale M, Lee E, Couto E, et al. Polymorphisms in hormone metabolism and growth factor genes and mammographic density in Norwegian postmenopausal hormone therapy users and non-users. *Breast Cancer Res* 2012; 14:R135.

11. Vachon CM, Scott CG, Faesching PA, et al. Common breast cancer susceptibility variants in LSP1 and RAD51L1 are associated with mammographic density measures that predict breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2012;21:1156–66.

12. Lindström S, Vachon CM, Li J, et al. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. *Nat Genet* 2011;43:185–7.

13. Stevens KN, Lindström S, Scott CG, et al. Identification of a novel percent mammographic density locus at 12q24. *Hum Mol Genet* 2012;21:3299–305. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22533574. Last accessed on May 29, 2012.

14. Satagopan JM, Verbel DA, Venkatraman ES, et al. Two-stage designs for gene-disease association studies. *Biometrics* 2002;58:163–70.

15. Satagopan JM, Venkatraman ES, Begg CB. Two-stage designs for gene-disease association studies with sample size constraints. *Biometrics* 2004;60:589–97.

16. Lope V, Pérez-Gómez B, Moreno MP, et al. Childhood factors associated with mammographic density in adult women. *Breast Cancer Res Treat* 2011;130:965–74.

17. Lope V, Pérez-Gómez B, Sánchez-Contador C, et al. Obstetric history and mammographic density: a population-based cross-sectional study in Spain (DDM-Spain). *Breast Cancer Res Treat* 2012;132:1137–46.

18. Cabanes A, Pastor-Barrilero R, García-López M, et al. Alcohol, tobacco, and mammographic density: a population-based study. *Breast Cancer Res Treat* 2011;129:135–47.

19. Garrido-Estapa M, Ruiz-Perales F, Miranda J, et al. Evaluation of mammographic density patterns: reproducibility and concordance among scales. *BMC Cancer* 2010;10:485.

20. Hoggart CJ, Whittaker JC, De Iorio M, et al. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet* 2008;4:e1000130.

21. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012. Available at: http://www.r-project.org/. Last accessed on May 30, 2012.

22. Lange K, Boehnke M. Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *Am J Med Genet* 1988;34:513–24.

23. Lange K, Weeks D, Boehnke M. Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* 1988;5:471–2.

24. Johnson AD, Handsaker RE, Pultz S, et al. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;24:2938–9. SNP Annotation and Proxy Search, SNAP. Broad Institute. 2012. Available at: http://www.broadinstitute.org/mpg/snap/lidsearch.php. Last accessed on November 20, 2012.

25. Guðbjartsson DF, Walters GB, Thorleifsson G, et al. Many sequence variants affecting diversity of adult human height. *Nat Genet* 2008;40:609–15.

26. Pollán M, Lope V, Miranda-García J, et al. Adult weight gain, fat distribution and mammographic density in Spanish pre- and post-menopausal women (DDM-Spain). *Breast Cancer Res Treat* 2012;134:823–38.

27. Assi V, Warwick J, Cuzick J, et al. Clinical and epidemiological issues in mammographic density. *Nat Rev Clin Oncol* 2012;9:33–40.

28. Sovio U, Bennett AJ, Millwood IY, et al. Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet* 2009;5:e1000409.

29. Lei S-F, Tan L-J, Liu X-G, et al. Genome-wide association study identifies two novel loci containing FLNB and SBFI genes underlying stature variation. *Hum Mol Genet* 2009;18:1601–9.

30. Okada Y, Kamatani Y, Takahashi A, et al. A genome-wide association study in 39 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum Mol Genet* 2010;19:2303–12.

31. Zhao J, Li M, Bradfield JP, et al. The role of height-associated loci identified in genome-wide association studies in the determination of pediatric stature. *BMC Med Genet* 2010;11:96.

32. Wishart MJ, Dixon JE. PTEN and myotubularin phosphatases: from 3-phosphoinositide dephosphorylation to disease. *Trends Cell Biol* 2002;12:579–85.

33. Ota T, Suzuki Y, Nishikawa T, et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* 2004;36:40–5.

34. Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov* 2012;2:40–1. cBio Cancer Genomics Portal. Computational Biology Center, Memorial Sloan-Kettering Cancer Center and i-Vis (Information Visualization) Research Group of the Computer Engineering Department, Bilkent University. 2012. Available at: http://www.cbioportal.org/public-portal/. Last accessed on November 26, 2012.

35. Lucci MA, Orlando R, Trizulji T, et al. Expression profile of tyrosine phosphorylated in HER2 breast cancer cells and tumors. *Cell Oncol* 2012;30:361–72.

36. McNally RS, Davis BK, Clements CM, et al. DJ-1 enhances cell survival through the binding of Ceranne, a negative regulator of NF-kappaB. *J Biol Chem* 2011;286:4098–106.

37. Pareja F, Ferraro DA, Rubin C, et al. Deubiquitination of EGFRT by Ceranne-1 contributes to cancer progression. *Oncogene* 2012;31:5499–608.

38. Gaudet MM, Kirchhoff T, Green T, et al. Common genetic variants and modifications of penetrance of BRCA2-associated breast cancer. *PLoS Genet* 2010;6:e1001183.

39. Turnbull C, Ahmed S, Morrison J, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 2010;42:504–7.

40. Couch FJ, Gaudet MM, Antoniou AC, et al. Common variants at the 19p13.1 and ZNF365 loci are associated with ER subtypes of breast cancer and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. *Cancer Epidemiol Biomarkers Prev* 2012;21:645–57.

41. García-Arenzana N, Navarrete-Muñoz EM, Peris M, et al. Diet quality and related factors among Spanish female participants in breast cancer screening programs. *Menopause* 2012;19:1121–9.

42. Li D, Lewinger JP, Gauderman WJ, et al. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol* 2011;35:790–9.

43. Harvey JA. Quantitative assessment of percent breast density: analog versus digital acquisition. *Technol Cancer Res Treat* 2004;3:611–16.

44. Pollán M, Asuncion N, Ederra M, et al. Mammographic density and risk of breast cancer according to tumor characteristics and mode of detection: a Spanish population-based case-control study. *Breast Cancer Res* 2013;15:R9.