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

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

Pablo Fernandez-Navarro,[1,2,*] Anna González-Neira,[3] Guillermo Pita,[3] Ramón Díaz-Uriarte,[4] Leticia Tais Moreno,[3] María Ederra,[2,5] Carmen Pedraz-Pingarrón,[6] Carmen Sánchez-Contador,[7] Jose Antonio Vázquez-Carrete,[8] Pilar Moreo,[9] Carmen Vidal,[10] Dolores Salas-Trejo,[11] Jennifer Stone,[12] Melissa C.Southey,[13,14] John L.Hopper,[15,16] Beatriz Pérez-Gómez,[1,2] Javier Benitez,[3,17,18] Marina Pollan,[1,2].

[1]Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid, 28029, 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, Health Promotion for women and childhood. General Directorate Public Health and Participation, Regional Authority of Health and Consumer Affairs, Balearic Islands, Spain.

[8]Galicia Breast Cancer Screening Programme, Regional Authority of Health, Galicia Regional Government, A Coruna, 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, University of Western Australia.

[13]Department of Pathology, University of Melbourne.

[14]Melbourne Collaborative Cohort Study (MCCS).

[15]School of Population Health, Centre for Molecular Environmental, Genetic and Analytic Epidemiology, University of Melbourne.

[16]School of Public Health, Seoul National University, 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.

[*] Corresponding author:

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

Novelty & Impact Statements

Our study is one of the largest conducted to date assessing the association between genetics and mammographic density, an intermediate phenotype for breast cancer, and covering pre and post-menopausal women. Using a two-stage genome-wide association analysis among Spanish women together with a replication analysis in women from an Australian study, we found evidence, not described before, that genetic variants rs11205277, rs11205303 in gene MTMR11 and rs67807996 in gene OTUD7B are associated with mammographic density.

Abstract

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 3351 Caucasian women aged 45-68 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 575374 SNPs in 239 pairs of women with extreme phenotypes and a verification stage for the 183 selected SNPs in the remaining sample (2873 women). Replication was conducted in 1786 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. 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;pvalue=$1.33x10^{-10}$). 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 *MTMR11*(OR=0.73;95%CI=0.66-0.80;pvalue=$2.64x10^{-11}$) and rs67807996 in gene *OTUD7B*(OR=0.72;95%CI=0.66-0.80;pvalue=$2.03x10^{-11}$). Our findings provide additional evidence on common genetic variations that may contribute to MD.

Keywords: DDM-Spain, GWAs, Mammographic density, MTMR11, OTUD7B

Introduction

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.[1,2] Twin and family studies have shown that genetic variation accounts for at least 60% of MD variability.[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.[5–10] A pooled cross-sectional analysis[11] of common breast cancer susceptibility variants in 14 independent loci found variants in *LSP1* (rs3817198) and *RAD51L1* (rs10483813) genes associated with mammographic density measures. Recently, two previous genome-wide association studies (GWAS)[12,13] have identified new genetic variants associated with MD, one in gene *ZNF365* (rs10995190) and other between genes *TBX5* and *TBX3* (rs1265507). However, these single nucleotide polymorphisms (SNPs) only account for a 0.5–1.3% of the MD variance, leaving much of the genetic contribution to MD unexplained.

In the present study, we have done a two-stage GWAS[14,15] inside a multicentre study of 3574 Spanish women attending BC screening (DDM-Spain "Determinants of Density in Mammographies in Spain"), together with a replication analysis of our best hit in 1786 women from the Australian Twins and Sisters Study of Mammographic Density.[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.[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 3574 women (range 497 to 536 per centre). Of the total sample, only Caucasian women (3351) were genotyped and included in the present 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 semiquantitative 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).[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 similiar 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 Additional file 1) from 599011 BeadChip markers only 575374 SNPs were informative in our sample. To select the best candidates SNPs associated with low/high MD, we used to strategies based on p-values and a Bayesian approach (see Statistical Analysis).

In the second stage, we genotyped 172 SNPs from the 183 SNPs selected in stage 1 (those with an Illumina Score >0.60, see Additional file 1) in the remaining 2873 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 GenomeStudio 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 inconsistences. Only 156 SNPs of the 172 SNPs genotyped in this second stage were analyzed after implementing a quality control (see Additional 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[11] 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 1786 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.[8]

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 socio-demographic information (ie. weight, height, smoking history and cessation of menstruation) (see details [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[8]).

Statistical Analysis. Two-stage GWAS (DDM-Spain).

In the first stage ("feature selection stage"), that 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[20], 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 Additional file 1). A total of 183 SNPs fulfilled these criteria.

The second stage used the remained 2873 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 4 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 (Additional file 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 Additional files 4 and 5. The genetic inflation factor for the stage 1 was 1.020.

These statistical analyses were performed using R software[21]; 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 assuming for each sibship that the residuals followed a multivariate normal distribution. Models were fitted using the statistical software package FISHER.[22,23]

Association analysis of Linkage disequilibrium region

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)[24] to identify these genetic variants. The inputs used in the proxy search function of SNAP are the following: "1000 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 (3351 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 stratus, age and BMI).

Results

Two-stage GWAS (DDM-Spain)

Table 1 shows the characteristics of the study participants (3351 Caucasian women). The majority (87% and 75% for stage 1 and 2 respectively) were postmenopausal, with a mean age of 57 vs 56 years old and a mean BMI of 26.10 vs 28.35 in stage 1 vs 2 respectively. The percentage of women with MD>50% and <10% in the whole sample were 23% and 25% respectively.

Table S1 in Additional file 6 shows the results of the association analysis for the 183 SNPs that fulfilled the established criteria in the stage 1. The majority of them were in chromosomes 1, 2 and 3 and their nominal p values not met the commonly used genome wide significance criterion of $p < 5 \times 10^{-8}$, being ranged between 4.72×10^{-6} and 9.58×10^{-4}.

Table S2 in Additional file 7 shows the results of the association analysis for the SNPs identified previously in the literature in this stage. Only rs10995190 showed a significant association (OR (per minor allele increase and likelihood of higher density category) =0.58 CI=0.40-0.85; p value= 4.45×10^{-3}).

The results of 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 value= 5.57 $x10^{-5}$), showing an Odds Ratio (OR) (per minor allele increase and likelihood of higher density category) of 0.77 (95%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.

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 & p values=0.07, 0.08 and 0.08 respectively) (see Table 4).

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 12579 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 values=2.64 $x10^{-11}$ and $2.03x10^{-11}$ 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 (3351 women) (see Table 3). The association between rs11205277 and MD (stage 1 and 2 samples combined) was also significant (p value= $1.33x10^{-10}$; OR (per minor allele increase and likelihood of higher density category)=0.74 CI=0.67-0.81). The genotype frequencies for these three SNPs by mammographic density category are shown in Table S3 (Additional 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 Additional 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 risk of MD associated with the variant rs11205277, located in an intergenic position at chromosome 1. No prior genome wide association study has shown an association between this SNP and MD.[12,13] The function of this genetic region is unclear. However, Gudbjartsson DF, et al.[25] found a high correlation of this SNP with height in Caucasian people, showing that the Histone class 2A, *MTMR11*, *SV2A*, and *SF3B4* genes are neighbouring the loci. In fact, BMI, a measurement closely related with height, is also associated with MD.[26,27] Another study[28] found that rs11205277 was associated with two stages of height growth, the peak height velocities in infancy (PHV1) and the puberty (PHV2). Moreover Shu-Feng,L. et al.[29] confirmed the association of this SNP with stature. Finally, Okada, Y. et al.[30] and Zao et al.[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 value=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 value=7.29×10^{-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=1.53×10^{-10}) was still significant after additional adjustment for height.

The replication analysis partiality 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.[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" (*MTMR11*) at 1q12-q21, which has been mentioned previously in relation with height.[25] This gene has a phosphatase activity[32,33] 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.[34] Moreover, Lucci MA. et al.[35] found *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 *OTUD7B* gene, making likely that it could affect gene expression by altering gene regulatory sequences. However the data in track "Chromatin State Segmentation by HMM from ENCODE/Broad" of USCS GenomeBrowser (http://genome.ucsc.edu/) don't 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-κB activity[36] and contributes to cancer progression by deubiquination of EGFR.[37] 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[12] and also in cancer susceptibility[38–40] 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.[5–10] In relation with the pooled cross-sectional analysis[11] 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 value= 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 Additional file 7). Finally, the two previous genome-wide association studies (GWAS)[12,13] 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 value= 4.45 $x10^{-3}$) in Stage 1 (see Additional file 7).

The differences in the study design, samples (Population-based versus 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 6-categories by an experienced reader. This classification has been associated with the risk of subsequent breast cancer in Spain.[44] In other studies, density was measured using a computer-assisted tool (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 papers describing an underestimation of percent of density using semi-authomatic 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 1662[13] to 4877.[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.[41] 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.[42]

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.[27,43] In our study, 3 of the 7 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.

Abbreviations

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

Acknowledgments

The authors declare no conflict of interest. The authors wish to 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-ISCIII)" (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 Santamariña, Montserrat Corujo and Ana Belén Fernández are with Galicia Breast Cancer Screening Programme, Regional Authority of Health, Galicia Regional Government, Galicia, Spain; María 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-García 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 Balearic Islands Breast Cancer Screening Programme, Health Promotion for women and childhood. General Directorate Public Health and Participation, Regional Authority of Health and Consumer Affairs, Balearic Islands, Spain.

This study was supported by two research grants from Spain's Health Research Fund (Fondo de Investigación Sanitaria) (FIS PI060386 and PIS09/01006); a Collaboration Agreement between Astra-Zeneca and the Carlos III Institute of Health (Instituto de Salud Carlos III) (EPY 1306/06); and a grant from the Spanish Federation of Breast Cancer Patients (FECMA 485 EPY 1170-10). Diaz-Uriarte R work was partially supported by project from the Spanish MINECO (BIO2009-12458)

References

1. Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, Paterson AD. Mammographic breast density as an intermediate phenotype for breast cancer. *Lancet Oncol* 2005;6:798–808.

2. Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Res* 2011;13:223.

3. Stone J, Dite GS, Gunasekara A, English DR, McCredie MRE, Giles GG, Cawson JN, Hegele RA, Chiarelli AM, Yaffe MJ, Boyd NF, Hopper JL. The heritability of mammographically dense and nondense breast tissue. *Cancer Epidemiol Biomarkers Prev* 2006;15:612–7.

4. Vachon CM, Sellers TA, Carlson EE, Cunningham JM, Hilker CA, Smalley RL, Schaid DJ, Kelemen LE, Couch FJ, Pankratz VS. Strong evidence of a genetic determinant for mammographic density, a major risk factor for breast cancer. *Cancer Res* 2007;67:8412–8.

5. Tamimi RM, Cox D, Kraft P, Colditz GA, Hankinson SE, Hunter DJ. Breast cancer susceptibility loci and mammographic density. *Breast Cancer Res* 2008;10:R66.

6. Woolcott CG, Maskarinec G, Haiman CA, Verheus M, Pagano IS, Le Marchand L, Henderson BE, Kolonel LN. 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, Van Den Berg D, Bernstein L, Ursin G. The role of established breast cancer susceptibility loci in mammographic density in young women. *Cancer Epidemiol Biomarkers Prev* 2008;17:258–60.

8. Odefrey F, Stone J, Gurrin LC, Byrnes GB, Apicella C, Dite GS, Cawson JN, Giles GG, Treloar SA, English DR, Hopper JL, Southey MC. 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, Santamariña C, Moreno MP, Vidal C, Miranda-García J, Ascunce N, Casanova F, Collado-García F, Herráez B, González-Neira A, Benítez J, et al. Association analysis between breast cancer 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;

10. Ellingjord-Dale M, Lee E, Couto E, Ozhand A, Qureshi SA, Hofvind S, Van Den Berg DJ, Akslen LA, Grotmol T, Ursin G. 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, Fasching PA, Hall P, Tamimi RM, Li J, Stone J, Apicella C, Odefrey F, Gierach GL, Jud SM, Heusinger K, 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, Varghese J, Thompson D, Warren R, Brown J, Leyland J, Audley T, Wareham NJ, Loos RJF, Paterson AD, 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, Lindstrom S, Scott CG, Thompson D, Sellers TA, Wang X, Wang A, Atkinson E, Rider DN, Eckel-Passow JE, Varghese JS, Audley T, et al. Identification of a novel percent mammographic density locus at 12q24. *Human Molecular Genetics [Internet]* 2012 [cited 2012 May 29];Available from: http://www.ncbi.nlm.nih.gov/pubmed/22532574

14. Satagopan JM, Verbel DA, Venkatraman ES, Offit KE, Begg CB. 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, Vidal C, Salas-Trejo D, Ascunce N, Román IG, Sánchez-Contador C, Santamariña MC, Carrete JAV, Collado-García F, Pedraz-Pingarrón C, 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, Santamariña MC, Moreo P, Vidal C, Laso MS, Ederra M, Pedraz-Pingarrón C, González-Román I, García-López M, Salas-Trejo D, 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-Barriuso R, García-López M, Pedraz-Pingarrón C, Sánchez-Contador C, Vázquez Carrete JA, Moreno MP, Vidal C, Salas D, Miranda-García J, Peris M, Moreo P, et al. Alcohol, tobacco, and mammographic density: a population-based study. *Breast Cancer Res Treat* 2011;129:135–47.

19. Garrido-Estepa M, Ruiz-Perales F, Miranda J, Ascunce N, González-Román I, Sánchez-Contador C, Santamariña C, Moreo P, Vidal C, Peris M, Moreno MP, Váquez-Carrete JA, 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, Balding DJ. 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. R Foundation for Statistical Computing [Internet]. Vienna, Austria: 2012 [cited 2012 May 30]. Available from: http://www.r-project.org/

22. Lange K, Boehnke M. Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *Am J Med Genet* 1983;14:513–24.

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

24. SNAP Proxy Search [Internet]. 2012 [cited 2012 Nov 20];Available from: http://www.broadinstitute.org/mpg/snap/ldsearch.php

25. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, Helgadottir A, Ingason A, 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, García M, Casanova F, Sánchez-Contador C, Santamariña C, Moreo P, Vidal C, Peris M, Moreno MP, Vázquez-Carrete JA, 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, Duffy SW. Clinical and epidemiological issues in mammographic density. *Nat Rev Clin Oncol* 2012;9:33–40.

28. Sovio U, Bennett AJ, Millwood IY, Molitor J, O'Reilly PF, Timpson NJ, Kaakinen M, Laitinen J, Haukka J, Pillas D, Tzoulaki I, Molitor J, 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, Wang L, Yan H, Guo Y-F, Liu Y-Z, Xiong D-H, Li J, Yang T-L, Chen X-D, Guo Y, et al. Genome-wide association study identifies two novel loci containing FLNB and SBF2 genes underlying stature variation. *Hum Mol Genet* 2009;18:1661–9.

30. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, Daigo Y, Yamamoto K, Kubo M, Nakamura Y, Kamatani N. A genome-wide association study in 19 633 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, Zhang H, Mentch FD, Wang K, Sleiman PM, Kim CE, Glessner JT, Hou C, Keating BJ, Thomas KA, 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, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* 2004;36:40–5.

34. cBio Cancer Genomics Portal [Internet]. 2012 [cited 2012 Nov 26];Available from: http://www.cbioportal.org/public-portal/

35. Lucci MA, Orlandi R, Triulzi T, Tagliabue E, Balsari A, Villa-Moruzzi E. Expression profile of tyrosine phosphatases in HER2 breast cancer cells and tumors. *Cell Oncol* 2010;32:361–72.

36. McNally RS, Davis BK, Clements CM, Accavitti-Loper MA, Mak TW, Ting JP-Y. DJ-1 enhances cell survival through the binding of Cezanne, a negative regulator of NF-kappaB. *J Biol Chem* 2011;286:4098–106.

37. Pareja F, Ferraro DA, Rubin C, Cohen-Dvashi H, Zhang F, Aulmann S, Ben-Chetrit N, Pines G, Navon R, Crosetto N, Köstler W, Carvalho S, et al. Deubiquitination of EGFR by Cezanne-1 contributes to cancer progression. *Oncogene* 2012;31:4599–608.

38. Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, Segrè AV, McGee K, McGuffog L, Kartsonaki C, Morrison J, Healey S, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. *PLoS Genet* 2010;6:e1001183.

39. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, 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, Ramus SJ, Kuchenbaecker KB, Soucy P, Beesley J, Chen X, Wang X, Kirchhoff T, McGuffog L, Barrowdale D, 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, Salas D, Ascunce N, Gonzalez I, Sánchez-Contador C, Santamariña C, Moreo P, Moreno MP, Carrete JAV, Collado-García F, 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, Murcray CE, Conti D. 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;3:611–6.

44. Pollan M, Ascunce N, Ederra M, Murillo A, Erdozain N, Ales-Martinez JE, Pastor-Barriuso R. 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.

Table legends

Table 1. Characteristics of the 3351 participants in the two stages of the DDM-Spain study.

Table 2. Results of the association analysis in stage 2 (2873 women and 156 SNPs).

Table 3. Results of the association analysis for the best hit and the two selected SNPs in the LD region including all women from DDM-Spain (3351 women, stages 1 and 2 combined)

Table 4. Results of the association analysis in the Replication analysis (Australian MD Twins and Sisters Study).

Figure legends

Figure 1. Forest plot of ORs for the association between MD and rs11205277 in stage 2 (2873 women). Forest plot of ORs and 95%CIs for the association analysis between MD and rs11205277 stratified by categories of age at mammography, BMI, menopausal status and recruiting center. Summary= OR and 95%CI of rs11205277 without stratification

Figure 2. Regional LD plot of rs11205277. Created by the web application of SNP Annotation and Proxy Search (SNAP).

Tables

Table 1. Characteristics of the 3351 participants in the two stages of the DDM-Spain study.

Variable		Stage	
		1	2
N		478	2873
Mean age at mammography (years)		57.31 (4.82)	55.88 (5.55)
Mean body mass index (kg/m^2)		26.10 (2.21)	28.35 (5.25)
Number of Post-menopausal women (%)		416 (87)	2155 (75)
Mammographic density categories	*1 (<10%)*	239 (50%)	585 (20%)
	2 (10-25%)	Nap	676 (24%)
	3 (25-50%)	Nap	1070 (37%)
	4 (>50%)	239 (50%)	542 (19%)

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

Table 3. Results of the association analysis for the best hit and the two selected SNPs in the LD region including all women from DDM-Spain (3351 women, stages 1 and 2 samples combined).

Polymorphism Locus	Chromosome	Position[a]	MAF[b]	HWE[c]	Alleles[d]	OR[e]	95%CI[f]	P-value[g]	N[h]	Brant test[i]
rs11205277	1	148159496	43	0.01	A/G	0.74	0.67-0.81	$1.33x10^{-10}$	3338 (0.4)	0.14
rs67807996	1	148261889	39	0.01	G/A	0.72	0.66-0.80	$2.03x10^{-11}$	3344 (0.2)	0.10
rs11205303	1	148173037	41	0.03	T/C	0.73	0.66-0.80	$2.64x10^{-11}$	3284 (2)	0.09

[a] Position= position of SNP (36.3 NCBI reference genome build); [b] MAF= Minor allele frequency (%); [c] HWE= p value of Hardy-Weinberg equilibrium test; [d] Alleles= major allele/minor allele; [e] OR= odds ratio, ordinal logistic regression model adjusted for recruitment centre, age at mammography, Body Mass Index (BMI) and menopausal status; [f] 95%CI= 95% confidence interval of OR; [g] P-value= p value for log-additive model; [h] N= number of individuals (percentage (%) of missing genotypes); [i] Brant test= p value of the likelihood-ratio test of the proportional odds assumption. The odds ratios are calculated taking the homozygous for the major allele as the reference category.

Table 3. Results of the association analysis for the best hit and the two selected SNPs in the LD region including all women from DDM-Spain (3351 women, stages 1 and 2 samples combined).

Polymorphism Locus	Chromosome	Position[a]	MAF[b]	HWE[c]	Alleles[d]	OR[e]	95%CI[f]	P-value[g]	N[h]	Brant test[i]
rs11205277	1	148159496	43	0.01	A/G	0.74	0.67-0.81	$1.33x10^{-10}$	3338 (0.4)	0.14
rs67807996	1	148261889	39	0.01	G/A	0.72	0.66-0.80	$2.03x10^{-11}$	3344 (0.2)	0.10
rs11205303	1	148173037	41	0.03	T/C	0.73	0.66-0.80	$2.64x10^{-11}$	3284 (2)	0.09

[a] Position= position of SNP (36.3 NCBI reference genome build); [b] MAF= Minor allele frequency (%); [c] HWE= p value of Hardy-Weinberg equilibrium test; [d] Alleles= major allele/minor allele; [e] OR= odds ratio, ordinal logistic regression model adjusted for recruitment centre, age at mammography, Body Mass Index (BMI) and menopausal status; [f] 95%CI= 95% confidence interval of OR; [g] P-value= p value for log-additive model; [h] N= number of individuals (percentage (%) of missing genotypes); [i] Brant test= p value of the likelihood-ratio test of the proportional odds assumption. The odds ratios are calculated taking the homozygous for the major allele as the reference category.

Table 4. Results of the association analysis in the Replication analysis (Australian MD Twins and Sisters Study).

Polymorphism Locus	Chromosome	Possition[a]	MAF[b]	HWE[c]	Alleles[d]	Genotype			N[h]	Outcome					
						Wild[e]	Heter[f]	Homo[g]		Absolute dense area			Percentage Dense Area		
										Estimate[i]	SE[j]	P-value[k]	Estimate[i]	SE[j]	P-value[l]
rs12650052	4	87357762	15	0.84	C/T	1293	452	41	1786	0,13	0.10	0.19	1.01	0.74	0.17
rs11205277	1	148159496	42	0.52	A/G	600	859	327	1786	-0.13	0.07	0.07	-0.74	0.54	0.17
rs6721728	2	55345558	0	0.93	A/G	1779	7	0	1786	-0,15	0.85	0.86	0.92	6.51	0.89
rs4466755	10	96622243	43	0.8	C/T	592	868	326	1786	-0,03	0.07	0.62	-0.42	0.53	0.42
rs4944707	11	86831745	48	0.11	A/G	470	925	391	1786	-0,11	0.07	0.17	-0.35	0.55	0.52
rs1402704	11	75617248	30	0.23	C/T	865	771	150	1786	-0,00	0.08	0.99	0.45	0.59	0.45
rs1622354	2	2935003	29	0.44	G/T	905	742	139	1786	0,00	0.08	0.96	0.19	0.60	0.75
rs1571526	13	97716489	35	0.98	T/C	766	807	213	1786	0,04	0.07	0.62	0.14	0.56	0.80
rs12607966	18	49921278	27	0.83	C/T	943	707	136	1786	0,14	0.08	0.08	0.95	0.60	0.12
rs17164879	5	122891860	21	0.81	C/T	1125	587	74	1786	-0,14	0.09	0.10	-0.10	0.66	0.88
rs1738822	6	47203703	21	0.03	G/A	1112	614	60	1786	0,05	0.09	0.54	0.38	0.67	0.58
rs7230021	18	49792489	28	0.13	G/A	929	700	157	1786	0,13	0.08	0.08	0.92	0.59	0.12
rs9861551	3	60664516	4	0.93	G/A	1639	144	3	1786	-0,07	0.17	0.70	-1.05	1.31	0.42
rs1780330	1	162969178	50	0	C/T	2	1784	0	1786	1,33	1.47	0.36	8.13	11.55	0.48

[a] Position= position of SNP (36.3 NCBI reference genome build); [b] MAF= Minor allele frequency (%); [c] HWE= p value of Hardy-Weinberg equilibrium test; [d] Alleles= major allele/minor allele; [e] Wild= number of participants with homozygous wild type; Heter= number of participants with heterozygous variant; Homo= number of participants with homozygous variant; N=total number of participants genotyped; Estimate= Estimated coefficient for the genetic variant in a linear regression, assuming an additive genetic model and adjusting for age and BMI; SE= Coefficient Standard Error; P-value= p value for the linear regression.

Additional files

Additional file 1: A Word document containing supplementary methods.

Additional file 2: A Word document containing a Figure (Figure S1) that shows the Eigentstrat method (Principal Component analysis) in stage 1 by MD in Boyd scale and by Recruitment center. Left plot shows first two components stratified by High and Low MD. Right plot shows both components stratified by recruitment center.

Additional file 3: A Word document containing a Figure (Figure S2) that shows the Eigentstrat method in stage 2 by mammographic density and by Recruitment center. Left plot shows both components stratified by mammographic density category. Right plot shows both components stratified by recruitment center.

Additional file 4: A Word document containing a Figure (Figure S3) that shows Quantile-quantile (Q-Q) plot in Stage 1 (478 women and 575374 SNPs). The observed p-values based on the GWAs performed in stage 1 are plotted against the expected distribution of p values under the null distribution.

Additional file 5: A Word document containing a Figure (Figure S4) that shows a Manhattan plot for the GWAs performed in Stage 1 (478 women and 575374 SNPs). The –log10(*p*) values are plotted against chromosomal base-pair position. The "Best hit" of our study is marked in red.

Additional file 6: A Word document containing a Table (Table S1) that shows the results of the association analysis of the 183 SNPs selected in stage 1.

Additional file 7: A Word document containing a Table (Table S2) that shows the results of the association analysis of known SNPs associated with MD (including women from Stage 1 of DDM-Spain (N=478)).

Additional file 8: A Word document containing a Table (Table S3) that shows the genotype frequencies for rs11205277, rs67807996 and rs11205303 by mammographic density category in the whole sample of DDM-Spain (3351 women).

Additional file 9: A Word document containing a Figure (Figure S5) that shows Forest plots of ORs and 95%CIs for the association analysis between MD and SNPs rs11205277, rs11205303 and rs67807996 in the whole sample (3351 women) stratified by categories of age at mammography, BMI, menopausal status and recruiting center.