Published in final edited form as:

J Med Genet. 2013 October ; 50(10): 666–673. doi:10.1136/jmedgenet-2013-101708.

Large-scale genotyping identifies a new locus at 22q13.2 associated with female breast size

Jingmei Li¹, Jia Nee Foo¹, Nils Schoof², Jajini S. Varghese^{3,4}, Pablo Fernandez-Navarro^{5,6}, Gretchen L. Gierach⁷, Swee Tian Quek⁸, Mikael Hartman⁹, Silje Nord¹⁰, Vessela N. Kristensen^{10,11}, Marina Pollán^{5,6}, Jonine D. Figueroa⁷, Deborah J. Thompson³, Yi Li¹, Chiea Chuen Khor¹, Keith Humphreys², Jianjun Liu^{1,9,*}, Kamila Czene^{2,*}, and Per Hall^{2,*}

¹Human Genetics, Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01 Singapore 138672, Singapore

²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, Stockholm 171 77, Sweden

³Centre for Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁴Plastic Surgery Unit, Royal Free Hospital, London

⁵Cancer and Environmental Epidemiology Unit, National Centre for Epidemiology, Carlos III Institute of Health, Avda Monforte de Lemos, 5, 28029 Madrid, Spain

⁶Consortium for Biomedical Research in Epidemiology & Public Health (*CIBER en Epidemiología y Salud Pública - CIBERESP*), Spain

⁷Hormonal and Reproductive Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

⁸Department of Diagnostic Radiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

⁹Saw Swee Hock School of Public Health, National University of Singapore and Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, MD3, 16 Medical Drive, Singapore 117597, Singapore

¹⁰Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, 0310 Oslo, Norway

Address for correspondence: Per Hall, MD, PhD, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, Stockholm 171 77, Sweden Tel: +46 8 524 86152 Fax: +46-8-31 49 75 Per.Hall@ki.se.

*Authors with equal contribution

Author Contributions: Overall study design was guided by KC, PH, JLi, JF, and KH.

Cohorts were supervised and phenotyped by KC, PH, NS, JLi, and JLi.

Analysis was performed by JLi, JV, PFN, GLG, STQ, MH, SN, VNK, MP, JDF, YL and DT.

The manuscript was written by JLi and JF.

All authors participated in critical review of the manuscript for intellectual content.

Disclosure: Nils Schoof is currently employed by Boehringer Ingelheim GmbH. Boehringer Ingelheim GmbH did not contribute any direct or indirect financing of this study.

Conflict Of Interest: None.

¹¹Department of Clinical Molecular Biology, Division of Medicine, Akershus University Hospital, 1478 Åhus, Norway

Abstract

Background—Individual differences in breast size are a conspicuous feature of variation in human females and have been associated with fecundity and advantage in selection of mates. To identify common variants that are associated with breast size, we conducted a large-scale genotyping association meta-analysis in 7,169 women of European descent across 3 independent sample collections with digital or screen film mammograms.

Methods—The samples consisted of the Swedish KARMA, LIBRO-1 and SASBAC studies genotyped on iCOGS, a custom illumina iSelect genotyping array comprising of 211,155 single nucleotide polymorphisms (SNPs) designed for replication and fine mapping of common and rare variants with relevance to breast, ovary and prostate cancer. Breast size of each subject was ascertained by measuring total breast area (mm²) on a mammogram.

Results—We confirm genome-wide significant associations at 8p11.23 (rs10086016, $P = 1.3 \times 10^{-14}$) and report a new locus at 22q13 (rs5995871, $P = 3.2 \times 10^{-8}$). The latter region contains the *MKL1* gene, which has been shown to impact endogenous estrogen-receptor α transcriptional activity and is recruited on estradiol-sensitive genes. We also replicated previous GWAS findings for breast size at four other loci.

Conclusion—A new locus at 22q13 may be associated with female breast size.

Keywords

Genome-wide association studies; population genetics; meta-analysis; breast size

Introduction

Individual differences in breast size are a conspicuous feature of variation in human females and have been associated with fecundity and advantage in selection of mates [1]. Breast size has also been implicated in other common conditions such as posture [2] and diabetes [3]. There is strong evidence of a genetic contribution to breast size. It has been estimated in a twin study that the heritability of bra cup size as a proxy for breast size is 56% [4]. Although this heritability is in part shared with body mass index (BMI), two thirds of the genes influencing breast size are expected to be unique [4].

Previous reports have relied on self-reported bra size as a proxy measure for breast size [3 5-6], but such measurements may not accurately reflect the actual breast size of the participants. Cup size labeling is not standardized; different brands of brassieres differ in their labeling of cup size for the same breast volume [7]. There is also the phenomenon known as “vanity sizing”, in which bra size is inflated. Such unreliability in bra sizing could possibly dilute any association and make generalizations of the results difficult [8]. In contrast, mammograms, which are X-ray images of breasts routinely used in screenings for the early detection of breast cancer, could possibly provide a more objective measure of breast size. There is also increasing interest in measuring the relationship between the

proportion of the breast composed of dense tissue and breast cancer risk. However, there has not been any prior attempt to use the total area of the breast outlined on a mammogram as a proxy for breast size. Here, we conducted a meta-analysis of three large-scale genotyping association studies (GWAS) on breast area ascertained from mammograms to confirm and identify novel loci associated with breast size.

Methods

Study participants

LIBRO-1 consisted of 5,125 breast cancer cases diagnosed between Jan 2001 – Dec 2008 from the Stockholm/Gotland area identified through the Stockholm breast cancer registry, and KARMA was comprised of 5,854 participants of the KARMA mammography screening study recruited between 2010 and 2011 from Helsingborg and Stockholm. A further 2,551 individuals were drawn from the SASBAC study, which comprised of women who were aged 50-74 and diagnosed with breast cancer in Sweden between 1993 and 1995 and population-based controls who were frequency matched by age to the cases.

From the parent studies above, mammogram collection was completed (as of May 1, 2012) for 3,420 cancer-free controls in the KARMA study and 2,243 breast cancer cases from the LIBRO-1 study which were included in our analyses. SASBAC included 732 breast cancer cases and 774 cancer-free controls with screen-film mammograms of the medio-lateral oblique view. In the final analysis, 7,169 samples were included. All sample collections were performed with Internal Review Board approval in Sweden with written informed consent from each study participant.

Mammograms and ascertainment of breast size

Breast size was ascertained as the breast area on a mammogram in mm². Mammograms of the cranial-caudal view in the KARMA and LIBRO-1 studies were automatically thresholded to find the breast edge and measured using ImageJ (Supplementary Figures 1 and 2). Mammograms of the medio-lateral oblique view in the SASBAC study were manually outlined by one reader to delineate the breast edge and exclude the pectoral muscle being before measured using Cumulus (Supplementary Figure 3). For the KARMA study which was comprised of cancer-free controls in a mammography screening program, breast size was taken as the average area measured of the left and right breasts. For breast cancer cases in the LIBRO-1 study, breast size was measured on the mammogram taken closest to diagnosis and on the side unaffected by cancer. Prediagnostic mammograms of the side unaffected by breast cancer were used for measurement in the SASBAC study for breast cancer cases; a random side was selected for the controls.

Genotyping and imputation

The SNP-based genotyping methodology has been previously described [9]. Genotyping for all three studies was conducted using a custom Illumina iSelect genotyping array (iCOGS), comprising 211,155 SNPs. Genotypes were called using Illumina's proprietary GenCall algorithm.

Standard GWAS QC criteria were applied: MAF $\geq 5\%$ with sample and SNP call rate $>95\%$ and Hardy–Weinberg equilibrium $P < 1 \times 10^{-5}$ using Plink (v1.07) [10]. We excluded closely related subjects based on identity-by-descent (IBD) probabilities. We estimated genome-wide IBD probabilities for each pair of subjects using the “—genome” option implemented in PLINK [10]. None of the individuals were found to have an IBD proportion > 0.4 . We also checked for subjects whose ancestries were estimated to be distinct from the other subjects by using PCA performed by EIGENSTRAT version 4.2 [11–12]. We visually inspected PCA plots for outliers in terms of ancestry from CEU clusters. Supplementary Figure 4 shows the projection of KARMA, LIBRO-1 and SASBAC samples onto the first two PCs of genetic variation. The spread for KARMA and LIBRO-1 is typical for samples drawn from populations of European ancestry [13]. As SASBAC was drawn from an earlier cohort with less influence from recent global migration trends, the sample appeared more homogeneous.

To select the number of PCs to adjust for population structure, we checked mammographic breast area for association with the PCs in each study [14]. We tested for PCs which were significantly associated with our phenotype of interest by treating them as covariates in our association analyses. Unlike logistic regression, adjusting for covariates associated with the trait in linear regression always improves the precision of the effect estimate by reducing the residual variance [15]. None of the top PCs were found to be associated with breast size in the three studies.

To increase resolution and coverage [16], subjects genotyped on the iCOGS genotyping array were imputed using the IMPUTE (v2.0) software based on the 1000 Genomes Project (Phase I integrated variant set release (v3), 19 Apr 2012) using multi-population reference panels [17]. The cleaned datasets were split by chromosomes, and then pre-phased separately for each chromosome using SHAPEIT v1.532 [18] prior to imputation [19]. Data was imputed separately for KARMA, LIBRO-1 and SASBAC.

We applied a series of post-imputation QC steps in order to eliminate unreliably imputed SNPs, aiming to filter out as many of these SNPs as possible while retaining a good proportion of nonsignificant SNPs. Post-imputation QC filtering was based on the IMPUTE-info score, which is associated with the imputed allele frequency estimate which ranges from 1, indicating high confidence, to 0 suggesting decreased confidence. SNPs with an IMPUTE-info score of < 0.8 and MAF $< 1\%$ were excluded.

Statistical analysis

We used the program SNPTTEST (v2.3.0) to test each SNP for the association with square-root transformed mammographic breast area in mm^2 for normality using a score test, analogous to the Cochran–Armitage trend test, but modified to cope with uncertainty in genotype calls. Phenotypes were mean centered and scaled to have variance 1. All analyses were adjusted for age at mammogram in years and BMI (kg/m^2). Due to the differences in breast cancer case-control status and type of mammogram, we derived a combined test statistic for each SNP by combining p-values and the direction of association for KARMA, LIBRO-1 and SASBAC, weighted by the sample size and the study-specific inflation factor, as previously described [20]. The numbers of SNPs available for meta-analysis from

KARMA, LIBRO-1 and SASBAC were 4,310,379, 4,359,979 and 4,299,570 respectively (Supplementary Table 1). The union of these three data sets (4,492,143 SNPs) was meta-analyzed using the program METAL (Mar 25 2011 release) [21]. METAL calculates a z -statistic for each marker summarizing the magnitude and direction of the effect relative to the reference allele in each sample and then calculates an overall Z -statistic and p value from the weighted average of the statistics. Genomic control correction was applied during the meta-analysis to adjust for ancestry. The total sample size comprises of 7,169 individuals. All association results were expressed relative to the forward strand of the reference genome.

To evaluate whether there is any genetic overlap between mammographic breast area and breast cancer risk, polygenic scores (based on mammographic breast area) were calculated according to the methods by Purcell *et al.* [22] to test the combined effect of multiple weak associations across the genome. The SASBAC data, which has screen film mammograms for both breast cancer cases and controls, were used. In order to identify polygenic effects due to independent SNPs in linkage equilibrium with one another, we performed the analysis using a pruned dataset (51,303 SNPs). Linkage equilibrium-based SNP pruning was performed using the `--indep-pairwise` function (window size in SNPs: 1,500; the number of SNPs to shift the window at each step: 150; r^2 threshold: 0.2) in Plink [10]. KARMA, the study with the most number of samples ($n=3,420$) was used as a reference dataset for computing polygenic scores. Ten sets of SNPs were generated from the mammographic breast area association results at P value thresholds of 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 and 1.0. Each subject was then scored using the sum of the number of reference alleles multiplied by the corresponding beta estimate from the mammographic breast area GWAS for the alleles in each of the SNP sets. Logistic regression was used compare the polygenic scores between breast cancer cases and controls. All models were adjusted for age at mammogram and BMI.

Estimation of the phenotypic variance explained by common variants for 1,206 subjects genotyped on both the iCOGS and Illumina HumanHap300 BeadChip was performed using the REML method described in Yang et al [23]. The estimate of variance explained for breast cancer was transformed from the observed scale to that on the underlying scale based on a specified prevalence rate of 0.0189.

Results

We performed a meta-analysis of three large-scale genotyping association studies (GWAS) to identify genetic variants that are associated with adult female breast size. Breast size is ascertained from the total breast area (mm^2) outlined on mammograms (Supplementary Figures 1-3). Table 1 summarizes the descriptive statistics and baseline demographic characteristics for the different studies involved.

Figures 1 and 2 show a Manhattan plot and a quantile-quantile (QQ) plot displaying the $\log_{10}$ transformed P value (y -axis) for each of the analyzed SNPs in the meta-GWAS under the additive model. The genomic inflation factor (λ) was 1.03, 1.01 and 1.03 for KARMA, LIBRO-1 and SASBAC (Supplementary Table 1), respectively, indicating that the extent to

which population stratification confounded the results was minimal. Only two SNPs showed genome-wide significant associations and consistent evidence of replication across all 3 datasets (rs10086016, $P = 1.3 \times 10^{-14}$, $\beta = -0.16$; $I^2 = 0$; $P_{\text{het}} = 0.54$ and rs5995871, $P = 3.2 \times 10^{-8}$, $\beta = -0.13$; $I^2 = 0$; $P_{\text{het}} = 0.94$, Table 2) under the additive model. Regional plots of the combined association results and recombination rates for 8p11.23 and 22q13.2 are shown in Figure 3.

The strongest evidence of association with mammographic breast area ($P = 1.3 \times 10^{-14}$) in the combined analysis was found at rs10086016, which maps to 8p11.23 at position 36,847,709. This SNP is in perfect linkage disequilibrium with a SNP (rs7816345) previously associated with self-reported bra size [5]. All associated SNPs within this locus were in linkage disequilibrium (LD) with rs10086016 and conditional analyses revealed no additional independent signals within a 500 kb window (Supplementary Figure 5).

We also identified a novel locus at 22q13.2 (rs5995871, genotyped SNP with smallest P-value) which reached genome-wide significance at $P = 3.2 \times 10^{-8}$, mapping within *MKLI*. Imputation of 1KGP SNPs yielded multiple variants associated with total mammographic breast area, with the lowest P-value observed at rs73169028 ($P = 1.22 \times 10^{-8}$). This SNP was in high LD with rs5995871 ($R^2 = 0.833$; $D' = 0.913$). In conditional regression analyses, rs5995871 was sufficient to account for all association signals within the locus (± 500 kb, Supplementary Figure 5). No heterogeneity in the beta estimates was observed for either locus across all three studies. A list of all genotyped SNPs which passed the suggestive threshold of $P < 1 \times 10^{-5}$ is given in Supplementary Table 2. Five other SNPs (rs4849887 and rs17625845 near *INHBB*, rs12173570 near *ESR1*, rs7089814 in *ZNF365*, rs12371778 near *PTHLH*) that have previously been reported to be associated with bra size were also found to be associated with total mammographic breast area in our study at $P < 0.05$ (Supplementary Table 3).

It was previously reported in Eriksson et al. [5] that another variant near the *MKLI* gene, rs73167017, showed suggestive association with bra size, although the P-value did not surpass genome-wide significance (5.47×10^{-7}). However, this variant was not significantly associated with total mammographic breast area in our data ($P = 0.1066$) and was in low LD ($R^2 = 0.062$; $D' = 0.756$) with rs5995871.

Allele frequency differences by breast size were observed among both European and Asian populations for the top hits in the 8p11.23 and 22q13.2 regions. Every additional copy of rs10086016(C) and rs5995871(G) was associated with smaller total mammographic breast area. The frequencies of these alleles were both lower for the European CEU population (each $\sim 12\%$), than for the Asian CHB and JPT populations on HapMap (ranging between 24.5% to 38.6%) (Supplementary Table 4).

Given that there is a public interest in knowing whether there is a link between breast size and breast cancer, we examined whether breast cancer susceptibility loci were associated with mammographic breast size. In the landmark breast cancer large-scale genotyping association analysis involving more than 100,000 women conducted by the Breast Cancer Association Consortium (BCAC), 67 known and novel breast cancer susceptibility loci were

reported [9]. Of the 13 breast cancer susceptibility loci associated (at $P < 0.05$) with mammographic breast area only six SNPs showed a consistent (the allele associated with large size being also associated with breast cancer risk) direction of association (Table 3 and Supplementary Table 5). In a lookup of our top genotyped SNPs for total mammographic breast area in the consortium data, rs10086016(C) and rs5995871(G) were found to be associated with overall breast cancer risk (odds ratio [95% confidence interval]) of 0.95 (0.93 to 0.97) and 1.12 (1.09 to 1.15), respectively. Both SNPs were associated with a decrease in total mammographic breast area. Whilst polygenic scores derived from the “top” SNPs for association with total mammographic breast area in the KARMA dataset were found to be significantly associated with the same trait in the SASBAC dataset for Pvalue thresholds between 0.1 and 1 ($P < 0.05$, Supplementary Figure 6), the smallest pvalue for association with breast cancer risk in the same dataset was 0.14.

We also evaluated the proportion of phenotypic variance that could possibly be explained by all SNPs present on the current genotyping array used. For a subset of the SASBAC study ($n = 1,206$), we had genetic data generated from both the customized iCOGS chip and the commercial Infinium HumanHap300 BeadChip. We estimated the proportion of phenotypic variance for mammographic breast area explained by all SNPs on each chip to be 0.31 (iCOGS, $SE = 0.16$) and 0.47 (Illumina, $SE = 0.25$), respectively (Supplementary Table 6). For breast cancer, the estimated proportion of phenotypic variance explained were 0.57 (iCOGS, $SE = 0.11$) and 0.32 (Illumina, $SE = 0.17$), respectively.

Discussion

Breast size is widely perceived to be influenced by genetic factors, but until recently, there has been no attempt to investigate which genes might be responsible for conferring variance to this phenotype. Eriksson et al. [5] performed a genome-wide association study (GWAS) of self-reported bra cup size, controlling for age, genetic ancestry, breast surgeries, pregnancy history and bra band size, in a cohort of more than 16,000 Caucasian women and found seven SNPs significantly associated with breast size at a genome-wide level. Using total mammographic breast area as an alternate method that is less prone to misclassification bias to measure breast size, we confirmed their genome-wide significant hit at 8p11.23 and identified a novel locus at 22q13.2. Together, the two SNPs explain between 0.6 to 1% of the variance for this phenotype, which is in keeping with most GWAS studies of traits and diseases assaying common genetic variants of low penetrance [20–24]. There was no evidence of epistasis between the two SNPs ($P > 0.05$).

Our most significant hit, located in the previously reported 8p11 locus, rs10086016, is in complete LD ($r^2 = 1.0$) with rs7816345, which was previously described. Based on newly available ENCODE data [25], rs10086016 may be associated with regulatory motif alterations that may influence transcriptional regulation in the nuclear factor I (NFI) gene family (Supplementary Figure 7). Unique NFI proteins have been reported to be expressed during lactation and involution, and hence the NFI gene family may play multiple important roles throughout mammary gland development [26].

The novel 22q13.2 locus encompasses the gene *MKLI* (megakaryoblastic leukemia [translocation] 1). The protein encoded by this gene interacts with the transcription factor myocardin, a key regulator of smooth muscle cell differentiation [27]. The encoded protein is predominantly located in the nucleus and may help transduce signals from the cytoskeleton to the nucleus [28-29]. In addition, *MKLI* has been shown to impact endogenous estrogen-receptor α transcriptional activity and is recruited on estradiol-sensitive genes [30]. We further looked up potential *cis* associations between rs5995871 and the expression levels of nearby (± 1 Mb) genes in 123 normal breast tissue samples in an unpublished expression trait quantitative loci (eQTL) dataset, and found the SNP to be significantly associated ($P < 0.05$) with the expression of the *ACO2*, *SGSM3*, *ST13* and *L3MBTL2* genes (Supplementary Figure 8).

It is possible that the allele frequency differences for the top hits in both the 8p11.23 and 22q13.2 regions are linked to ethnic differences in breast size. Asian women tend to have smaller breasts on average when compared to Caucasian women [31]. In this study, we observed that the effect alleles of the two common variants which were found to be associated with smaller total mammographic breast area to be more common amongst the Chinese and Japanese HapMap populations compared to Europeans. However, it should be noted that since these SNPs explain only about 1% of the variance for breast size, a relatively slight difference in allelic frequency may explain only a small fraction of the actual phenotype.

Study findings on a potential link between breast size and breast cancer are mixed. Some studies have found that larger breast sizes are associated with increased risk of cancer, while others have found no link between breast size and cancer risk [32-35]. We expanded the analysis carried out by Eriksson et al. [5] to include 67 novel and known breast cancer susceptibility SNPs reported in a landmark paper on a mega-consortium effort (COGS) to identify breast cancer susceptibility variants [9] and found 13 breast cancer SNPs to be associated with breast size at $P < 0.05$, suggesting that there is a significant overlap among SNPs for breast size and breast cancer risk ($P = 2.91 \times 10^{-7}$ in a two-tailed test of population proportion where the alternate hypothesis is that the true proportion is not equal to 5% as expected by chance). It is however, notable that the effects of 6 of the 13 SNPs were in opposite directions for breast size and breast cancer risk. Although we have found evidence that there is a connection between genetic determinants of breast size and breast cancer risk, our results show that larger breast size are not necessarily associated with an increase breast cancer risk. On the iCOGS chip which has been customized to study genetic variants predisposing to breast cancer, we also saw no evidence of a shared polygenic component between mammographic breast area and the disease, suggesting that caution should be exercised when interpreting any positive correlation between these two phenotypes.

The strengths of this study lie in the objectiveness and reproducibility of the measurement of breast size. However, given our validation of the results from Eriksson et al [5], we concur that there is merit in using bra size as a convenient proxy measurement for breast size. Nonetheless, it is not uncommon that different countries use different bra sizing methods; thus, area measurements based on mammograms are likely more generalizable. Most of the women in our analyses were also peri- or postmenopausal and past child-bearing age, which

minimizes potential confounding from changes in breast size due to pregnancy or lactation. Further assessment of the heritability of mammographic breast area or its correlation with bra size will provide further insight into the validity of these trait measurements for genetic and epidemiological studies of breast size.

A possible limitation is that the estimated proportion of phenotypic variance explained by all SNPs on the iCOGS chip is only ~67% that of the commercial genome-wide SNP chips, thus limiting the number of novel discoveries made in our study. We were able to replicate five other SNPs in four loci previously found to be associated with mammographic breast area with $P < 0.05$. The eight known breast size SNPs to date were found to collectively explain between 1 to 2.4% of the variance for this phenotype in our three data sets. We do, however, acknowledge the relatively lower power compared to Eriksson's study [5] on bra size with more than twice as many samples. Further studies on larger samples with mammographic breast area and better coverage will allow for the identification of additional loci associated with breast size.

Another potential limitation of our study could be the mixture of screen-film and digital mammograms, as well as mammograms of different views (i.e. CC or MLO). Breast area tends to be higher when assessed from digital mammograms rather than from films, in part due to the better delineation of the breast edge on a digital mammogram [36]. The mean breast area measured from mammograms of the MLO view is also higher than that of the CC view, due to the inclusion of the tissue in the upper quadrant of the breast. However, the type and view of mammograms were uniform within each individual study, and we took into account the fact that effect sizes may be different by carrying out the meta-analysis using a p-value based analysis, weighted for sample size and direction of effect [21].

In summary, we confirmed a locus at 8p11.23 and identified a novel locus at 22q13.2 for breast size using an objective and reliable phenotype. We found no clear evidence that breast cancer susceptibility SNPs were associated with breast size. Further studies are required to provide a more comprehensive understanding of the genetics of breast size by combining GWAS with large samples drawn from diverse ethnic populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

KARMA and LIBRO-1 were supported by Märit and Hans Rausing's Initiative Against Breast Cancer. SASBAC was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. All three studies are genotyped as part of the Collaborative Oncological Gene-environment Study (COGS), which is supported by the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS). Genotyping of the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer program and the Ministry of Economic Development, Innovation and Export Trade of Quebec (grant PSR-SIIRI-701). BCAC is funded by Cancer Research UK (C1287/A10118 and C1287/A12014). Combining the GWAS data was supported in part by the US National Institutes of Health (NIH) Cancer Post-Cancer GWAS initiative grant 1 U19 CA 148065-01 (DRIVE, part of the GAME-ON initiative). KC was financed by the Swedish Cancer Society (5128-B07-01PAF). KH was supported by the Swedish Research Council (523-2006-972) and the Swedish E Science Research Centre. This work was also supported, in

part, by the Intramural Research Program of the U.S. National Cancer Institute, Department of Health and Human Services, USA. We wish to thank Aki Tuulainen, Kimberley Sio Kim Chua, Sander Canisius, Grethe I.G. Alnæs, Torben Luders and Lodewyk Wessels for their help in data collection and analysis.

References

1. Buggio L, Vercellini P, Somigliana E, Vigano P, Frattaruolo MP, Fedele L. "You are so beautiful" *: Behind women's attractiveness towards the biology of reproduction: a narrative review. *Gynecol Endocrinol*. 2012 published Online First: Epub Date. 10.3109/09513590.2012.662545
2. Findikcioglu K, Findikcioglu F, Ozmen S, Guclu T. The impact of breast size on the vertebral column: a radiologic study. *Aesthetic Plast Surg*. 2007; 31(1):23–7. published Online First: Epub Date. 10.1007/s00266-006-0178-5 [PubMed: 17205252]
3. Ray JG, Mohllajee AP, van Dam RM, Michels KB. Breast size and risk of type 2 diabetes mellitus. *Cmaj*. 2008; 178(3):289–95. published Online First: Epub Date. 10.1503/cmaj.071086 [PubMed: 18227448]
4. Wade TD, Zhu G, Martin NG. Body mass index and breast size in women: same or different genes? *Twin research and human genetics : the official journal of the International Society for Twin Studies*. 2010; 13(5):450–4. published Online First: Epub Date. 10.1375/twin.13.5.450 [PubMed: 20874466]
5. Eriksson N, Benton GM, Do CB, Kiefer AK, Mountain JL, Hinds DA, Francke U, Tung JY. Genetic variants associated with breast size also influence breast cancer risk. *BMC Medical Genetics*. 2012; (13):53. [PubMed: 22747683]
6. Oo M, Myint Z, Sakakibara T, Kasai Y. Relationship between brassiere cup size and shoulder-neck pain in women. *Open Orthop J*. 2012; 6:140–2. published Online First: Epub Date. 10.2174/1874325001206010140 [PubMed: 22509231]
7. Pechter EA. A new method for determining bra size and predicting postaugmentation breast size. *Plast Reconstr Surg*. 1998; 102(4):1259–65. [PubMed: 9734454]
8. Sulem P, Gudbjartsson DF, Geller F, Prokopenko I, Feenstra B, Aben KK, Franke B, den Heijer M, Kovacs P, Stumvoll M, Magi R, Yanek LR, Becker LC, Boyd HA, Stacey SN, Walters GB, Jonasdottir A, Thorleifsson G, Holm H, Gudjonsson SA, Rafnar T, Bjornsdottir G, Becker DM, Melbye M, Kong A, Tonjes A, Thorgeirsson T, Thorsteinsdottir U, Kiemeny LA, Stefansson K. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. *Hum Mol Genet*. 2011; 20(10):2071–7. ddr086 [pii]. published Online First: Epub Date. 10.1093/hmg/ddr086 [PubMed: 21357676]
9. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Breast Ovarian Cancer Susceptibility C; Fletcher O, Peto J, Gibson L, Dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Czene K, Irwanto A, Liu J, Waisfisz Q, Meijers-Heijboer H, Adank M, Hereditary B Ovarian Cancer Research Group N; van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Muller-Myhsok B, Lichtner P, Hopper JL, Southey MC, Makalic E, Schmidt DF, Uitterlinden AG, Hofman A, Hunter DJ, Chanock SJ, Vincent D, Bacot F, Tessier DC, Canisius S, Wessels LF, Haiman CA, Shah M, Luben R, Brown J, Luccarini C, Schoof N, Humphreys K, Li J, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang X, Vachon C, Stevens KN, Lambrechts D, Moisse M, Paridaens R, Christiaens MR, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken Z, Aaltonen K, Heikkinen T, Broeks A, Veer LJ, van der Schoot CE, Guenel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Zamora MP, Perez JJ, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MW, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Henderson BE, Schumacher F, Le Marchand L, Andrulis IL, Knight JA, Glendon G, Mulligan AM, kConFab I; stralian Ovarian Cancer Study G; Lindblom A, Margolin S, Hooning MJ, Hollestelle A, van den Ouweland AM, Jager A, Bui QM, Stone J, Dite GS, Apicella C, Tsimiklis H, Giles GG, Severi G, Baglietto L, Fasching PA, Haeblerle L, Ekici AB, Beckmann MW, Brenner H, Muller H, Arndt V, Stegmaier C, Swerdlow A, Ashworth A, Orr N, Jones M, Figueroa J, Lissowska J, Brinton L, Goldberg MS, Labreche F, Dumont M, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Bruning T, Network G. Radice P, Peterlongo P, Manoukian S, Bonanni B, Devilee P, Tollenaar RA, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma

VM, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dork T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo KY, Noh DY, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng CC, Van Den Berg D, Stram DO, Shu XO, Lu W, Gao YT, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng JH, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsang P, Shen CY, Hsiung CN, Wu PE, Ding SL, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Blot WJ, Signorello LB, Cai Q, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Simard J, Garcia-Closas M, Pharoah PD, Chenevix-Trench G, Dunning AM, Benitez J, Easton DF. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature genetics*. 2013; 45(4):353–61. published Online First: Epub Date. 10.1038/ng.2563 [PubMed: 23535729]

10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. S0002-9297(07)61352-4 [pii]. published Online First: Epub Date. 10.1086/519795 [PubMed: 17701901]
11. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38(8):904–9. ng1847 [pii]. published Online First: Epub Date. 10.1038/ng1847 [PubMed: 16862161]
12. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2(12):e190. 06-PLGE-RA-0101R3 [pii]. published Online First: Epub Date. 10.1371/journal.pgen.0020190 [PubMed: 17194218]
13. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatos E, D'Alfonso S, Blackburn H, Martinelli Boneschi F, Liddle J, Harbo HF, Perez ML, Spurkland A, Waller MJ, Mycko MP, Ricketts M, Comabella M, Hammond N, Kockum I, McCann OT, Ban M, Whittaker P, Kempainen A, Weston P, Hawkins C, Widaa S, Zajicek J, Dronov S, Robertson N, Bumpstead SJ, Barcellos LF, Ravindrarajah R, Abraham R, Alfredsson L, Ardlie K, Aubin C, Baker A, Baker K, Baranzini SE, Bergamaschi L, Bergamaschi R, Bernstein A, Berthele A, Boggild M, Bradfield JP, Brassat D, Broadley SA, Buck D, Butzkueven H, Capra R, Carroll WM, Cavalla P, Celius EG, Cepok S, Chiavacci R, Clerget-Darpoux F, Clysters K, Comi G, Cossburn M, Cournu-Rebeix I, Cox MB, Cozen W, Cree BA, Cross AH, Cusi D, Daly MJ, Davis E, de Bakker PI, Debouverie M, D'Hooghe MB, Dixon K, Dobosi R, Dubois B, Ellinghaus D, Elovaara I, Esposito F, Fontenille C, Foote S, Franke A, Galimberti D, Ghezzi A, Glessner J, Gomez R, Gout O, Graham C, Grant SF, Guerini FR, Hakonarson H, Hall P, Hamsten A, Hartung HP, Heard RN, Heath S, Hobart J, Hoshi M, Infante-Duarte C, Ingram G, Ingram W, Islam T, Jagodic M, Kabesch M, Kermode AG, Kilpatrick TJ, Kim C, Klopp N, Koivisto K, Larsson M, Lathrop M, Lechner-Scott JS, Leone MA, Leppa V, Liljedahl U, Bomfim IL, Lincoln RR, Link J, Liu J, Lorentzen AR, Lupoli S, Macciardi F, Mack T, Marriott M, Martinelli V, Mason D, McCauley JL, Mentch F, Mero IL, Mihalova T, Montalban X, Mottershead J, Myhr KM, Naldi P, Ollier W, Page A, Palotie A, Pelletier J, Piccio L, Pickersgill T, Piehl F, Pobywajlo S, Quach HL, Ramsay PP, Reunanen M, Reynolds R, Rioux JD, Rodegher M, Roesner S, Rubio JP, Ruckert IM, Salvetti M, Salvi E, Santaniello A, Schaefer CA, Schreiber S, Schulze C, Scott RJ, Sellemjerg F, Selmaj KW, Sexton D, Shen L, Simms-Acuna B, Skidmore S, Sleiman PM, Smestad C, Sorensen PS, Sondergaard HB, Stankovich J, Strange RC, Sulonen AM, Sundqvist E, Syvanen AC, Taddeo F, Taylor B, Blackwell JM, Tienari P, Bramon E, Tourbah A, Brown MA, Tronczynska E, Casas JP, Tubridy N, Corvin A, Vickery J, Jankowski J, Villoslada P, Markus HS, Wang K, Mathew CG, Wason J, Palmer CN, Wichmann HE, Plomin R, Willoughby E, Rautanen A, Winkelmann J, Wittig M, Trembath RC, Yaouanq J, Viswanathan AC, Zhang H, Wood NW, Zuvich R, Deloukas P, Langford C, Duncanson A, Oksenberg JR, Pericak-Vance MA, Haines JL, Olsson T, Hillert J, Ivinson AJ, De Jager PL, Peltonen L, Stewart GJ, Hafler DA, Hauser SL, McVean G, Donnelly P, Compston A. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011; 476(7359):214–9. nature10251 [pii]. published Online First: Epub Date. 10.1038/nature10251 [PubMed: 21833088]

14. Peloso GM, Lunetta KL. Choice of population structure informative principal components for adjustment in a case-control study. *BMC Genet.* 2011; 12:64. 1471-2156-12-64 [pii]. published Online First: Epub Date. 10.1186/1471-2156-12-64 [PubMed: 21771328]
15. Robinson L, Jewell N. Some Surprising Results about Covariate Adjustment in Logistic Regression Models. *International Statistical Review.* 1991; 59(2):227–40.
16. Zhang W, Dolan ME. Impact of the 1000 genomes project on the next wave of pharmacogenomic discovery. *Pharmacogenomics.* 2010; 11(2):249–56. published Online First: Epub Date. 10.2217/pgs.09.173 [PubMed: 20136363]
17. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009; 5(6):e1000529. published Online First: Epub Date. 10.1371/journal.pgen.1000529 [PubMed: 19543373]
18. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2012; 9(2):179–81. nmeth.1785 [pii]. published Online First: Epub Date. 10.1038/nmeth.1785 [PubMed: 22138821]
19. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics.* 2012
20. Lindstrom S, Vachon CM, Li J, Varghese J, Thompson D, Warren R, Brown J, Leyland J, Audley T, Wareham NJ, Loos RJ, Paterson AD, Rommens J, Waggott D, Martin LJ, Scott CG, Pankratz VS, Hankinson SE, Hazra A, Hunter DJ, Hopper JL, Southey MC, Chanock SJ, Silva Idos S, Liu J, Eriksson L, Couch FJ, Stone J, Apicella C, Czene K, Kraft P, Hall P, Easton DF, Boyd NF, Tamimi RM. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. *Nat Genet.* 2011; 43(3):185–7. ng.760 [pii]. published Online First: Epub Date. 10.1038/ng.760 [PubMed: 21278746]
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26(17):2190–1. published Online First: Epub Date. 10.1093/bioinformatics/btq340 [PubMed: 20616382]
22. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009; 460(7256):748–52. nature08185 [pii]. published Online First: Epub Date. 10.1038/nature08185 [PubMed: 19571811]
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011; 88(1):76–82. S0002-9297(10)00598-7 [pii]. published Online First: Epub Date. 10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]
24. Cornes BK, Khor CC, Nongpiur ME, Xu L, Tay WT, Zheng Y, Lavanya R, Li Y, Wu R, Sim X, Wang YX, Chen P, Teo YY, Chia KS, Seielstad M, Liu J, Hibberd ML, Cheng CY, Saw SM, Tai ES, Jonas JB, Vithana EN, Wong TY, Aung T. Identification of four novel variants that influence central corneal thickness in multi-ethnic Asian populations. *Hum Mol Genet.* 2012; 21(2):437–45. ddr463 [pii]. published Online First: Epub Date. 10.1093/hmg/ddr463 [PubMed: 21984434]
25. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012; 40(Database issue):D930–4. gkr917 [pii]. published Online First: Epub Date. 10.1093/nar/gkr917 [PubMed: 22064851]
26. Murtagh J, Martin F, Gronostajski RM. The Nuclear Factor I (NFI) gene family in mammary gland development and function. *J Mammary Gland Biol Neoplasia.* 2003; 8(2):241–54. [PubMed: 14635798]
27. Du KL, Chen M, Li J, Lepore JJ, Mericko P, Parmacek MS. Megakaryoblastic leukemia factor-1 transduces cytoskeletal signals and induces smooth muscle cell differentiation from undifferentiated embryonic stem cells. *J Biol Chem.* 2004; 279(17):17578–86. M400961200 [pii]. published Online First: Epub Date. 10.1074/jbc.M400961200 [PubMed: 14970199]
28. Ma Z, Morris SW, Valentine V, Li M, Herbrick JA, Cui X, Bouman D, Li Y, Mehta PK, Nizetic D, Kaneko Y, Chan GC, Chan LC, Squire J, Scherer SW, Hitzler JK. Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. *Nat Genet.* 2001; 28(3):220–1. 90054 [pii]. published Online First: Epub Date. 10.1038/90054 [PubMed: 11431691]
29. Sasazuki T, Sawada T, Sakon S, Kitamura T, Kishi T, Okazaki T, Katano M, Tanaka M, Watanabe M, Yagita H, Okumura K, Nakano H. Identification of a novel transcriptional activator, BSAC, by

- a functional cloning to inhibit tumor necrosis factor-induced cell death. *J Biol Chem.* 2002; 277(32):28853–60. M203190200 [pii]. published Online First: Epub Date. 10.1074/jbc.M203190200 [PubMed: 12019265]
30. Huet G, Merot Y, Percevault F, Tiffocche C, Arnal JF, Boujrad N, Pakdel F, Metivier R, Flouriot G. Repression of the estrogen receptor- α transcriptional activity by the Rho/megakaryoblastic leukemia 1 signaling pathway. *J Biol Chem.* 2009; 284(49):33729–39. M109.045534 [pii]. published Online First: Epub Date. 10.1074/jbc.M109.045534 [PubMed: 19826002]
31. Forbes GB, Frederick DA. The UCLA Body Project II: Breast and Body Dissatisfaction among African, Asian, European and Hispanic American College Women. *Sex Roles.* 2008; (58):449–57.
32. Tavani A, Pregnolato A, La Vecchia C, Negri E, Favero A, Franceschi S. Breast size and breast cancer risk. *Eur J Cancer Prev.* 1996; 5(5):337–42. [PubMed: 8972252]
33. Egan KM, Newcomb PA, Titus-Ernstoff L, Trentham-Dietz A, Baron JA, Willett WC, Stampfer MJ, Trichopoulos D. The relation of breast size to breast cancer risk in postmenopausal women (United States). *Cancer Causes Control.* 1999; 10(2):115–8. [PubMed: 10231159]
34. Kusano AS, Trichopoulos D, Terry KL, Chen WY, Willett WC, Michels KB. A prospective study of breast size and premenopausal breast cancer incidence. *Int J Cancer.* 2006; 118(8):2031–4. published Online First: Epub Date. 10.1002/ijc.21588 [PubMed: 16284954]
35. Thurfjell E, Hsieh CC, Lipworth L, Ekblom A, Adami HO, Trichopoulos D. Breast size and mammographic pattern in relation to breast cancer risk. *Eur J Cancer Prev.* 1996; 5(1):37–41. [PubMed: 8664807]
36. Assi V, Warwick J, Cuzick J, Duffy SW. Clinical and epidemiological issues in mammographic density. *Nat Rev Clin Oncol.* 2012; 9(1):33–40. published Online First: Epub Date. 10.1038/nrclinonc.2011.173 [PubMed: 22143145]

Figure 1.

Manhattan plot combining association results from 4,492,143 SNPs from the genome-wide meta-analysis based on mammograms collected as of May 15th 2012. SNPs genotyped on the iCOGS genotyping array were imputed using the IMPUTE (v2.0) software based on the 1000 Genomes Project (Phase I integrated variant set release (v3), 19 Apr 2012) using multi-population reference panels. Sample-specific association results were computed with SNPTEST (v2.3), meta-analyses were performed in METAL (Mar 25 2011 release). Meta-analysis results are plotted as $-\log_{10}$ P-values (y-axis) against physical chromosomal location (x-axis). Black triangles indicate genotyped SNPs $P < 1.0 \times 10^{-5}$. The plot was generated in R using “qqman”.

Figure 2.
Quantile-quantile (QQ) plot for the combined association results ($\lambda_{GC} = 1.02$)

Figure 3.

Regional plots of the genome-wide meta-analysis association results and recombination rates for the 8p11.23 and 22q13.2 regions (a–b). Association results of both genotyped (squares) and imputed (circles) SNPs in the combined samples and recombination rates within the loci at 8p11.23 (a) and 22q13.2 (b). For each plot, $-\log_{10} P$ values (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The top genotyped SNP in each combined analysis is shown as a purple square and is labeled by its rsID. Color intensity of each symbol reflects the extent of LD with the top genotyped SNP. Genetic recombination rates, estimated using 1000 Genomes EUR samples (Nov 2010), are shown with a light blue line. Physical positions are based on NCBI Build 37 of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association.

Table 1

Descriptive statistics for the different studies

Study	n	Type	View	Breast size	Age at mammogram	BMI
KARMA	3,420	Digital, processed	CC	137.2 (58.0)	54.2 (9.4)	25.2 (4.3)
LIBRO-1	2,243	Digitized screen-film	CC	124.4 (43.3)	58.4 (8.8)	25.3 (4.1)
SASBAC	1,506	Digitized screen-film	MLO	171.6 (54.3)	62.5 (6.4)	25.6 (3.9)
7,169						

Counts refer to subjects who passed all quality control filters

Breast size in cm², mean (s.d.). 1 cm² = 100 mm².

Age in years; mean (s.d.).

Body mass index in kg/m²; mean (s.d.).

CC: Cranial-caudal; MLO: medio-lateral oblique

Summary results for common variants at 8p11.23 and 22q13 associated with female breast size. Beta estimates and standard errors are reported on the square-root transformed scale.

Table 2

Region	SNP ¹	Type	Study	MAF ²	Beta	SE	P value
8p11.23	rs10086016(C)	Genotyped	KARMA	0.15	-0.11	0.02	4.0×10^{-6}
			LIBRO-1	0.15	-0.18	0.03	4.4×10^{-7}
			SASBAC	0.15	-0.16	0.04	9.3×10^{-5}
						1.3×10^{-14}	
(P _{het} = 0.54, I ² = 0%)							
22q13.2	rs5995871(G)	Genotyped	KARMA	0.11	-0.11	0.03	1.0×10^{-4}
			LIBRO-1	0.13	-0.11	0.04	3.9×10^{-3}
			SASBAC	0.11	-0.13	0.04	4.4×10^{-3}
						3.2×10^{-8}	
(P _{het} = 0.94, I ² = 0%)							

¹ SNP identifier. Letter in parentheses indicated the effect allele based on the forward strand.

² Minor allele frequency

Table 3
Associations between previously reported breast cancer associated SNPs and breast size ($P<0.05$)¹

Locus	SNP ²	Chr ³	Position ⁴	Br+Ca effect ⁵	Direction ⁶	P-value ⁷	HetISq ⁸	HetPVal ⁹
MKL1	rs6001930 (T)*	22	40876234	-	+++	1.4×10^{-6}	0	0.92
	rs17356907 (A)*	12	96027759	+	---	9.6×10^{-6}	0	0.49
NTN4	rs3757318 (A)	6	1.52E+08	+	+++	8.2×10^{-5}	0	0.53
ESR1	rs2046210 (A)	6	1.52E+08	+	+++	7.7×10^{-4}	0	0.52
PTHLH	rs10771399 (A)*	12	28155080	+	+++	1.3×10^{-3}	0	0.82
ANKRD16	rs2380205 (T)	10	5886734	-	+++	5.9×10^{-3}	0	0.42
8q24	rs13281615 (A)*	8	1.28E+08	-	---	0.01071	30.2	0.24
-	rs12422552 (C)	12	14413931	+	+++	0.01706	7.2	0.34
CCDC88C	rs941764 (A)*	14	91841069	-	+++	0.01781	0	0.67
FGFR2	rs2981579 (A)	10	1.23E+08	+	---	0.02048	22.9	0.27
ZNF365	rs10995190 (A)	10	64278682	-	+-	0.02049	61.6	0.07
-	rs4849887 (T)	2	1.21E+08	-	+++	0.03402	1.4	0.36
TOX3	rs3803662 (A)	16	52586341	+	+++	0.03751	0	0.65

¹ Associations between all 67 previously reported breast cancer associated SNPs and breast size are listed in Supplementary Table 6

² Published SNP showing the strongest association with breast cancer risk at that locus in European populations. Letter in parentheses indicated the effect allele based on the forward strand. An asterisk indicates effect directions are with respect to the major allele.

³ Chromosome

⁴ Build 37 Position

⁵ Direction of effect allele shown to be associated with breast cancer risk

⁶ Direction - summary of effect direction for each study, with one '+' or '-' per study (KARMA, LIBRO-1 and SASBAC, in that order)

⁷ P-value for association between SNP and breast size in this report; for $I2 < 25$, fixed effect model. For $I2 \geq 25$, random effects model.

⁸ HetISq - I^2 statistic which measures heterogeneity on scale of 0-100%

⁹ P-value for heterogeneity statistic