Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170

A full list of authors and affiliations appears at the end of the article.

Abstract

We analyzed 3,872 common genetic variants across the ESR1 locus (encoding estrogen receptor α) in 118,816 subjects from three international consortia. We found evidence for at least five independent causal variants, each associated with different phenotype sets, including estrogen receptor (ER+ or ER−) and human ERBB2 (HER2+ or HER2−) tumor subtypes, mammographic density and tumor grade. The best candidate causal variants for ER− tumors lie in four separate enhancer elements, and their risk alleles reduce expression of ESR1, RMND1 and CCDC170, whereas the risk alleles of the strongest candidates for the remaining independent causal variant disrupt a silencer element and putatively increase ESR1 and RMND1 expression.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

Correspondence should be addressed to A.M.D. (amd24@medschl.cam.ac.uk), D.F.E. (dfe20@medschl.cam.ac.uk) or S.L.E. (stacey.edwards@qimrberghofer.edu.au).

129 These authors contributed equally to this work.

198 These authors jointly supervised this work.

Accession codes. The relevant SNP genotype data underpinning these analyses can be accessed by applying to the BCAC and CIMBA consortia (see URLs).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS


COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.
SNPs at 6q25.1 have been reported to be associated with breast cancer susceptibility in genome-wide association studies (GWAS) in women of Chinese\(^1\) and European\(^2\) ancestry. Subsequent analyses have demonstrated that SNPs in the same region are associated with breast cancer risk for \textit{BRCA1} mutation carriers\(^3\) and mammographic density\(^4\), a strong breast cancer risk factor. Thus far, however, attempts to identify the candidate causal variant(s) underlying the associations have been inconclusive\(^3, 5, 6\). Here we report fine-scale mapping and comprehensive analysis of the genotype-phenotype associations in this region, using dense genotyping and imputed data from the custom-designed iCOGS array, in 118,816 subjects from three consortia: the Breast Cancer Association Consortium (BCAC), the Consortium of Investigators of Modifiers of \textit{BRCA1} and \textit{BRCA2} (CIMBA) and the Markers of Density Consortium (MODE). We additionally demonstrate, through functional analyses, the likely modes of action of the strongest candidate causal variants.

**RESULTS**

**Genetic epidemiological studies**

We successfully genotyped 902 SNPs across a 1-Mb region containing \textit{ESR1} in 50 case-control studies from populations of European (89,050 participants) and Asian (12,893 participants) ancestry in BCAC, together with 15,252 \textit{BRCA1} mutation carriers in CIMBA. Mammographic density measures were available for 6,979 women from the BCAC studies and an additional 1,621 women from the MODE Consortium, who had also been genotyped using the iCOGS array. Subsequently, the genotypes of additional variants with minor allele frequency (MAF) >2% were imputed in all European-ancestry participants, using data from the 1000 Genomes Project as a reference. In total, data from 3,872 genotyped or imputed (imputation info score >0.3) SNPs were analyzed. Results for all SNPs associated with overall breast cancer risk (\(P < 1 \times 10^{-4}\)) are presented in Supplementary Table 1. Manhattan plots of the associations of these 3,872 SNPs with the main phenotypes are shown in Figure 1.

**Conditional analyses**

All genotyped and imputed SNPs displaying evidence of association with overall breast cancer risk in women of European ancestry (\(P < 1 \times 10^{-4}\)) were initially included in forward stepwise logistic regression models for \textit{ER}^- and \textit{ER}^+ breast tumor risk. The most parsimonious models (Online Methods) included four SNPs for \textit{ER}^- breast cancer and four SNPs for \textit{ER}^+ breast cancer, with three SNPs being common to both models. In each model, all selected SNPs fell into a subset of five bins of correlated SNPs (\(r^2 >0.8\)). Stepwise regression models were independently fitted to breast cancer risk in the CIMBA \textit{BRCA1} mutation carriers and to mammographic density (measured as mammographic dense area; see the Online Methods for full details). For the \textit{BRCA1} mutation carriers and for mammographic dense areas, the SNPs in the best fitting models also fell within a subset of the five originally defined bins. For further analyses, we selected the directly genotyped SNP that was most significantly associated with the predominant phenotype for that bin. Regression analyses were repeated using just these five SNPs, with each representing an independent signal\(^7\). Results are presented in Table 1. Additionally, in the BCAC studies, we
were able to examine SNP associations with risks of HER2 (HER2+ and HER2−) and progesterone receptor (PR+ and PR−) tumor subtypes and with tumor grade at diagnosis. There were weak but detectable correlations between the representative SNPs for signals 1–4 (Table 1 and Supplementary Table 2). We therefore modeled the associations with each SNP conditional on the other four; these conditional risk estimates and significance levels are also presented in Table 1. At conditional significance levels of $P < 1 \times 10^{-3}$, four of the lead SNPs (signals 1, 2, 4 and 5) were independently associated with risk of developing ER− breast cancer (Table 1). Another, partially overlapping, set of four SNPs (signals 1–3 and 5) was associated with ER+ tumor risk (Table 2 and Supplementary Table 3), and another subset of SNPs (signals 1–4) was associated with breast cancer risk in BRCA1 mutation carriers (Table 1). The per-allele odds ratios were higher for ER− than for ER+ disease for three lead SNPs (signals 1, 2 and 5), whereas representative SNPs for signal 3 displayed smaller effects of similar magnitude on risk for ER− and ER+ tumors. Mammographic dense area was associated with representative SNPs from signal 2 and less strongly with those from signal 1 (Table 1). We additionally carried out a meta-analysis of the SNP associations with breast cancer risk for CIMBA BRCA1 mutation carriers and risk of ER− tumors in BCAC. We anticipated that this analysis would increase statistical power to detect ER− risk signals, and, indeed, it did strengthen the evidence for association of SNPs representing signals 1–4 but not signal 5, which showed no association with breast cancer risk in BRCA1 mutation carriers (Table 1).

**Tumor subtype and grade analyses**

We next explored the associations of each signal with specific tumor subtype combinations and with tumor grade (Fig. 1f, Table 2 and Supplementary Tables 3–5). The representative SNPs at two signals (3 and 5) were strongly associated with high-grade disease, after adjusting for ER status ($P < 1 \times 10^{-3}$; Table 2 (bottom line) and Supplementary Table 5). Among ER− tumors, three signals (1, 2 and 4) were associated with triple-negative (ER−PR−HER2−) and high-grade tumors, as well as the rarer ER−PR−HER2+ subtype, with similar odds ratios (Table 2 and Supplementary Tables 3 and 5). However, signal 5 was more strongly associated with ER−PR−HER2+ disease (odds ratio (OR) = 1.24, 95% confidence interval (CI) = 1.12–1.37; $P = 2.4 \times 10^{-5}$; Table 2) than with the triple-negative subtype (OR = 1.08, 95% CI = 1.01–1.15; $P = 0.016$; Table 2, case-only $P = 0.021$; Supplementary Table 5), consistent with the lack of association for breast cancer in BRCA1 mutation carriers, in whom tumors are predominantly triple negative8.

**Haplotype analysis**

We next explored the combined effects of the same five signal-representative genotyped SNPs (Supplementary Table 6). Haplotype-specific effects were consistent with additive effects of the individual signal-representative SNPs. In particular, haplotype 22221 (all minor alleles except for signal 5; frequency = 0.005) was associated with the largest increased risks of both ER+ (OR = 1.38, 95% CI = 1.11–1.71; $P = 3.3 \times 10^{-3}$) and ER− (OR = 2.34, 95% CI = 1.76–3.10; $P = 3.5 \times 10^{-3}$) tumors; this group includes the triple-negative (ER−PR−HER2−) tumor subtype (detected via the meta-analysis of BCAC subjects with ER− tumors and CIMBA BRCA1 mutation carriers; $P = 8 \times 10^{-10}$). Haplotype 22111 (frequency = 0.02) was associated with the highest risk of HER2+ tumors (OR = 1.5, 95% CI = 1.21–
1.87; \( P = 3 \times 10^{-4} \)) and with mammographic dense area (\( \beta \) coefficient = 0.45, 95% CI = 0.20 to 0.69; \( P = 3 \times 10^{-4} \)).

**Associations in Asian-ancestry studies**

We examined the associations of the five signal-representative SNPs in the nine Asian-ancestry studies in BCAC (Supplementary Table 7). All five displayed allelic associations in the same direction as in Europeans, with overlapping confidence intervals, consistent with the hypothesis that the same candidate causal variants determine risk in both populations.

**Determining the candidate SNPs within each signal**

To identify the potential causal variants to be taken forward for functional analysis, we determined the most significant SNP association within each signal and then calculated the likelihood ratio of every other SNP relative to that SNP. We assumed that SNPs with a likelihood of <1:100 (ref. 9) in comparison with the most significant SNP for each signal could be excluded from consideration as potentially causative variants. On the basis of the assumption that, within a given signal, the same variant(s) would be driving all observed phenotype associations, we derived the list of most likely causal SNPs for each signal. We used the results from one of two analyses to define the list of potentially causal SNPs for each signal: the meta-analysis of BCAC subjects with ER\(^-\) disease and CIMBA \( BRCA1 \) mutation carriers for signals 1, 2 and 4, which were most strongly associated in this analysis, and overall breast cancer risk in BCAC for signals 3 and 5. These lists of unexcluded variants are presented in Table 3 and are highlighted in Supplementary Table 1.

In signal 1, the most strongly associated variant was rs2046210 (the original Asian GWAS hit\(^1,10 \)), with nine other variants (likelihood ratios <100:1, \( r^2 \geq 0.89 \) with rs2046210; spanning 151,935,539–151,954,127) remaining as strong causal candidates. In signal 2, the best causal candidate was SNP rs12173570, with two other candidates remaining (likelihood ratios <100:1, \( r^2 \geq 0.75 \) with rs12173570; spanning 151,955,914–151,958,815). The European GWAS SNP, rs3757318l (ref. 2), is most strongly correlated with rs12173570 (\( r^2 > 0.45 \)). In signal 3, the best causal candidate was rs851984, with three other candidates remaining (likelihood ratios <100:1, \( r^2 = 0.99 \); in two \( ESR1 \) introns spanning 152,020,390–152,024,985). In signal 4, the top candidate was rs9918437, and two other candidates spanned another segment of an \( ESR1 \) intron at 152,055,978–152,072,718 (approximately 30 kb telomeric of signal 3; likelihood ratios <100:1, \( r^2 > 0.81 \) with rs9918437). In signal 5, the strongest candidate causal SNP was rs2747652 (also the representative SNP for signal 5 in Table 1), and there were five other candidates (likelihood ratios <100:1, \( r^2 > 0.97 \) with rs2747652; spanning 152,432,902–152,440,522) in the intergenic region between \( ESR1 \) and \( SYNE1 \). Across the five signals, we were able to exclude all but 26 of the original 3,872 variants from being potentially causal.

**Local gene expression analyses**

We used four techniques to assess associations between candidate causal variants (or available proxy SNPs) in the five signals and local gene expression. (i) ER protein expression, measured by immunohis-tochemistry in normal breast tissue samples from 150 postmenopausal donors, identified a significant correlation of the risk alleles of signal 1
SNPs with reduced ER levels (Fig. 2a and Supplementary Figs. 1 and 2). (ii) ESR1 expression in breast tumors and adjacent normal breast tissue from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) study was compared relative to signal-representative SNP allele (Fig. 2b and Supplementary Table 8). In patients with ER− tumors, risk allele carriers had lower median ESR1 expression in normal, tumor-adjacent tissue than homozygotes for the protective allele at signals 1, 4 and 5, although none of the differences were statistically significant. By contrast, in patients with ER+ tumors, risk allele carriers had higher median ESR1 expression in normal, tumor-adjacent tissue than homozygotes for the protective allele at signals 1, 3 and 5. (iii) Allele-specific expression (ASE) analysis, using RNA sequencing (RNA-seq) data from breast tumor samples and SNP array genotype data from The Cancer Genome Atlas (TCGA)11, showed allelic imbalances in ESR1 expression among heterozygotes for proxy SNPs in signals 1–3 (Fig. 2c and Supplementary Table 9). Similar imbalances in CCDC170 expression were detected among heterozygotes for signal 2 SNP rs9397437 and in RMND1 expression with signal 3 SNP rs851983 (Supplementary Table 9). Such allelic imbalances indicate that risk alleles at these signals are associated with expression differences in local genes, but they do not indicate the directions of association. (iv) Expression quantitative trait locus (eQTL) analysis using the Gene-Tissue Expression (GTEx) database identified a significant association for SNPs in signal 3 with CCDC170 expression in normal breast tissues (Supplementary Table 10). We also performed cis-eQTL analyses on the 12 flanking genes in 135 normal breast tissue samples from the METABRIC study; however, no additional associations were detected (Supplementary Table 11).

Bioinformatic and chromatin analyses

Analysis of cis enhancer–gene interactions using PreSTIGE12 showed evidence of multiple regulatory elements coinciding with signals 1–3 in ER+ MCF-7 breast cancer cells (Fig. 3a and Supplementary Fig. 3). A ‘super-enhancer’, associated with high levels of acetylation of histone H3 at lysine 27 (H3K27ac), was also identified in MCF-7 cells and encompasses the top risk-associated SNPs in these three signals (Fig. 3a and Supplementary Fig. 3)13. This super-enhancer was most readily detectable in MCF-7 cells and was not observed in other breast cancer cell lines, normal mammary epithelial cells or other tissues analyzed (Supplementary Fig. 4). Chromatin conformation capture (3C) experiments demonstrated that elements within signals 1 and 2 physically interacted with the promoters of ESR1, RMND1-ARMT1 and CCDC170 in MCF-7 and T-47D cells (Fig. 3b and Supplementary Fig. 5a,b). Furthermore, we detected interactions between signals 3–5 and ESR1 and/or RMND1-ARMT1 promoters (Fig. 3c,d and Supplementary Fig. 5c,d). The majority of these interactions were restricted to MCF-7 and T-47D cells (ER+ breast cancer cell lines), but the RMND1-ARMT1 interactions were also detected in either Bre-80 or MCF10A cells (ER− ‘normal’ breast cell lines; Fig. 3b–d and Supplementary Fig. 5b–d). The 3C-identified interactions for each signal are summarized in Supplementary Table 12.

Prioritizing candidate SNPs for functional assays

We applied a combination of in silico and in vitro analyses to prioritize candidate causal SNPs for functional follow-up, using previous observations that common cancer susceptibility alleles are enriched in cis-regulatory elements14–16. First, Table 3 showed that
19 of the 26 top candidates overlapped DNase I–sensitive sites and were associated with enhancer-enriched histone marks such as dimethylation of histone H3 at lysine 4 (H3K4me2) and H3K27ac in MCF-7 and HMEC breast cells, indicative of putative regulatory elements (PREs) (Supplementary Fig. 6). In electromobility shift assays (EMSAs), 11 of these 19 SNPs altered the binding affinity of transcription factors in vitro (Supplementary Fig. 7). Of these, seven fell within promoter-specific long-range interactions identified by 3C (Fig. 3 and Supplementary Fig. 5). The 7 SNPs prioritized for further detailed analyses included 2 of 10 remaining candidates in signal 1 (rs7763637 and rs6557160), 1 of 3 candidates in signal 2 (rs17081533), 2 of 4 candidates in signal 3 (rs851982 and rs851983), 1 of 3 candidates in signal 4 (rs1361024) and 1 of 6 candidates in signal 5 (rs910416) (Supplementary Table 12).

Luciferase reporter assays

The regulatory capabilities of the PREs overlapping each signal and the effects of the seven prioritized candidate SNPs were examined in luciferase reporter assays in the ER\(^+\) MCF-7 and BT-474 and the ER\(^-\) Bre-80 breast cell lines. PRE constructs containing the reference alleles of the prioritized SNPs for signals 1, 2, 4 and 5 significantly increased associated target gene promoter activity when cloned in either direction, indicating that they act as orientation-independent transcriptional enhancers. In contrast, a PRE containing the reference alleles of the signal 3 candidates ablated target gene promoter activity, but only when cloned in the forward direction, suggesting that this region acts as an orientation-dependent silencer (Fig. 4 and Supplementary Figs. 8–10). Notably, inclusion of the minor (risk) alleles of individual candidate SNPs in signals 1, 2 and 5 (rs6557160, rs17081533 and rs910416) significantly reduced ESR1 and RMND1 promoter activity but had no effect on the ARMT1 or CCDC170 promoters. However, inclusion of the signal 1 haplotype significantly decreased ESR1, RMND1 and CCDC170 promoter activity (Fig. 4 and Supplementary Figs. 8 and 9). Inclusion of the individual minor (risk) allele of signal 4 SNP rs1361024 or signal 3 SNP rs851983 in the respective constructs had no additional effects. In contrast, inclusion of the signal 3 minor (risk) allele of rs851982 or the haplotype construct increased ESR1 promoter activity in ER\(^+\) MCF-7 and BT-474 cells and RMND1 promoter activity in all three cell lines (Fig. 4, Supplementary Figs. 8 and 9, and Supplementary Table 12).

Transcription factor binding analyses

We used both bioinformatic analyses and functional studies to examine DNA-protein interactions for the seven prioritized SNPs. In silico prediction tools including intragenomic replicates (IGR)\(^{17}\), HaploReg\(^{18}\) and Alibaba2 (ref. 19) predicted that all seven SNPs alter transcription factor binding (Supplementary Fig. 11 and Supplementary Table 13).

Competition with known transcription factor binding sites suggested the identity of bound proteins for four of the prioritized SNPs, including GATA3 binding to the minor (risk) allele of signal 3 SNP rs851982 and CTCF binding to the minor allele of a second signal 3 candidate, rs851983, as well as the common (protective) allele of signal 4 candidate rs1361024 and MYC binding to the common allele of signal 5 candidate rs910416 (Supplementary Fig. 12 and Supplementary Table 12). Additional well-established breast
cell transcription factors, such as ER itself and FOXA1, were also assessed but did not display competitive binding to any prioritized SNP sites (Supplementary Fig. 13). Chromatin immunoprecipitation (ChIP) confirmed enrichment of GATA3 binding to DNA overlapping signal 3 candidate rs851982, but no difference between the alleles, and confirmed CTCF binding to the region overlapping signal 4 candidate rs1361024 in BT-474 cells (Fig. 5a and Supplementary Fig. 14). CTCF also bound to the region encompassing signal 3 candidate rs851983 (Fig. 5a, Supplementary Fig. 14 and Supplementary Table 12). CTCF mediates long-range chromatin looping; therefore, to assess the potential impact of signal 4 candidate rs1361024 and signal 3 candidate rs851983 on chromatin interactions, we performed allele-specific 3C in heterozygous cell lines. Sequence profiles indicated that the protective G allele of signal 4 candidate rs1361024 increases looping between this enhancer and the ESR1 and RMND1 promoters (Fig. 5b and Supplementary Fig. 15a). We found no evidence for allele-specific looping between the silencer overlapping signal 3 and local gene promoters (Supplementary Fig. 15b).

**DISCUSSION**

The fine-scale mapping, bioinformatic and functional analyses presented here provide evidence for the existence of at least five different genetic variants, each with a direct effect on breast cancer risk in Europeans, findings also supported by the limited available data in Asian populations. These variants are distributed upstream, within introns and downstream of ESR1, each in a region, which we have demonstrated via reporter assays, is regulatory for ESR1. Some may additionally regulate other local genes, such as RMND1, ARMT1 and CCDC170, previously reported to be co-regulated with ESR1 (ref. 20). Of note, the four sites more strongly associated with risks of ER− than ER+ tumors (signals 1, 2, 4 and 5) all overlap enhancer regions, and our evidence indicates that the minor (risk) alleles of candidate causal variants, within each of these enhancers, act to reduce expression of ESR1, RMND1 and CCDC170. In contrast, signal 3—which is associated with smaller but equal risks of developing both ER− and ER+ tumors—overlaps a putative gene silencer, and the risk alleles of the candidate causal variants here increase ESR1 and RMND1 expression. Furthermore, we have demonstrated altered binding of looping factor CTCF to candidate causal SNPs in signals 3 and 4, with evidence that the risk allele of signal 4 candidate rs1361024 abrogates binding and reduces chromatin looping between this enhancer element and the promoters of ESR1 and RMND1. We also provided evidence that signal 5 candidate rs910416 may display allele-specific binding of MYC.

Notably, the previously unrecognized signal 5 candidates, downstream of ESR1, significantly increase the risk of developing ER−PR−HER2+ tumors (a specific subtype shown to be more responsive to the drug trastuzumab) in contrast to the triple-negative (ER−PR−HER2−) tumor subtype, which has already been reported to be associated with other signals at 6q25 as well as 19p13 (ref. 21) and 5p15 (TERT)22. We also found evidence that the candidate causal variants at signals 3 and 5 predispose to aggressive, high-grade breast cancer, independently of ER status.

Mammographic density adjusted for age and body mass index (BMI), which describes the variation in epithelial and stromal tissue on a mammogram, is one of the strongest known
risk factors for breast cancer and has been shown to have a shared genetic basis with breast cancer, mediated through a large number of common variants. Associations between \( ESR1 \) SNPs and mammographic density have previously been reported, but, in this detailed analysis, only signal 2 was significantly associated with mammographic dense area \((P = 1.7 \times 10^{-5})\), although signal 1 also showed some evidence of an effect in the conditional analysis \((P = 0.017)\). Although adjusting the breast cancer analysis of signal 2 for mammographic dense area produced some attenuation of the associated effect, the lead SNP remained significantly associated with breast cancer risk \((\text{unconditional } OR = 1.30, 95\% \text{ CI} = 1.13–1.49; P = 0.00024; \text{OR conditional on dense area} = 1.24, 95\% \text{ CI} = 1.08–1.43; P = 0.0025)\), suggesting either that the mechanism by which the signal 2 candidate causal variant affects breast cancer risk is not mediated through mammographic density or, alternatively, that dense area, as measured here, is unable to capture the association with breast composition that is most relevant to risk. This phenomenon, whereby the association with risk appears to be partially independent of mammographic density, has also been observed for the 10q21.2 breast cancer locus.

SNPs in the \( ESR1 \) region have previously been reported to be associated with bone mineral density. These include SNPs within signal 1 \((rs6930633, r^2 = 0.73 \text{ with } rs3757322)\) and signal 3 \((rs2982575, r^2 = 0.57 \text{ with } rs851984)\), although the SNP with the most significant reported association with bone density measures, \(rs4870044\), was not associated with breast cancer risk \((P > 1 \times 10^{-4})\) in our analysis nor correlated with any signal-representative SNPs \((r^2 < 0.06)\). Similarly, SNP \(rs6933669\), recently reported as associated with age at menarche, is uncorrelated with these five signals \((r^2 < 0.02)\) and was not associated with breast cancer \((P = 0.1)\). Thus, although there is a known relationship between age at menarche and breast cancer risk, these phenotypes do not appear to share candidate causal variants in this region.

Our findings help address the question of the role of \(\text{ER} \alpha\) in establishing breast cancer. Notably, the candidate causal SNPs identified here all increase risks of both \(\text{ER}^+\) and \(\text{ER}^-\) tumor subtypes by varying degrees. \(\text{ER} \alpha\) is a ligand-activated transcription factor that mediates the effect of estrogen through altering gene expression, and the links between estrogen, \(\text{ER} \alpha\) and \(\text{ER}^+\) breast cancer are well documented, with adjuvant endocrine therapy considered standard treatment for \(\text{ER}^+\), early-stage breast cancer. Other studies have also reported 6q25 associations with \(\text{ER}^-\) subtypes, but the mechanisms by which \(\text{ER}^-\) tumors develop are still debated. There is speculation that \(\text{ER}^-\) tumors may arise from \(\text{ER}^+\) precursors by potentially reversible mechanisms, and our findings may lend support to this hypothesis. However, several recent studies have indicated that most tumors in \(\text{BRCA}1\) mutation carriers arise from \(\text{ER}^-\) luminal progenitor cells; thus, estrogen may be working indirectly through paracrine regulation in the mammary epithelium, possibly stimulating the Notch or epidermal growth factor receptor (EGFR) signaling pathways of adjacent \(\text{ER}^+\) cells. Our analyses unexpectedly suggested that, whereas signals 1–4 increased risks of all \(\text{ER}^-\) tumor subtypes, the signal 5 candidate causal variant increased risks of \(\text{ER}^-\text{HER2}^+\) breast cancer subtypes but not of triple-negative tumor development or of tumors in \(\text{BRCA}1\) mutation carriers (Table 1). This further complicates present understanding and underlines the need for further studies to address this issue.
Collectively, our evidence supports a hypothesis that \textit{ESR1} is the major target gene of the enhancer and silencer elements in which we have identified candidate causal variants. In addition to \textit{ESR1}, we provide evidence that the regions overlapping signals 1–4 cooperatively regulate \textit{RMND1}, raising the possibility that candidate causal SNPs act by altering both \textit{ESR1} and \textit{RMND1} expression. \textit{RMND1} (required for meiotic nuclear division 1; \textit{C6orf96}) has not been well characterized but is reported to localize to mitochondria and be involved in mitochondrial translation\textsuperscript{33}. We additionally identified enhancer activity and chromatin interactions with two other genes, \textit{ARMT1} and \textit{CCDC170}, but the actions of the candidate causal SNPs on these genes remain unclear. \textit{ARMT1} encodes Armt1, a protein carboxyl methyltransferase that targets PCNA and differentially regulates cancer cell survival in response to DNA damage\textsuperscript{34}. Nothing is known about the function of \textit{CCDC170} (coiled-coil domain–containing protein 170), but recurrent \textit{ESR1-CCDC170} rearrangements have been characterized in an aggressive subset of ER\textsuperscript{+} breast cancers\textsuperscript{35}. A recent study also showed that higher \textit{CCDC170} expression correlated with ER negativity, highly proliferative features and worse clinical outcomes\textsuperscript{36}. There are some data to suggest that these genes may cooperatively contribute to the increased proliferative capacity of ER\textsuperscript{+} tumors\textsuperscript{20}, and it is tempting to speculate that these may be additional target genes for the candidate causal variants at a subset of the five signals identified here and perhaps responsible for their differential phenotype associations. A greater understanding of these genes may also provide novel targets for breast cancer prevention or therapies.

**URLs**


**ONLINE METHODS**

**Study populations and genotyping**

Epidemiological data were obtained from three separate consortia that had all conducted genotyping using the iCOGS array, a custom array comprising approximately 200,000 SNPs. (i) Data on overall breast cancer risk, tumor subtypes and grade came from 50 breast cancer case-control studies participating in BCAC; these comprised 41 studies from populations of European ancestry and 9 studies from populations of East Asian ancestry\textsuperscript{3}. Details of the participating studies, genotype calling and quality control are given elsewhere\textsuperscript{3}. After quality control exclusions, we analyzed data from 46,451 cases and 42,599 controls of European ancestry and 6,269 cases and 6,624 controls of Asian ancestry. A further 23 SNPs were directly genotyped in two case-control studies (CCHS and SEARCH). The ER status of the primary tumor was available for 34,539 European and 4,972 Asian...
cases; of these, the tumor was ER− for 7,465 (22%) European and 1,610 (32%) Asian cases. (ii) Data on BRCA1 mutation carriers were obtained through CIMBA. Eligibility is restricted to females 18 years or older with pathogenic mutations in BRCA1 or BRCA2. The majority of the participants were sampled through cancer genetics clinics, including some related participants. Fifty-one studies from 25 countries contributed data on BRCA1 mutation carriers who were genotyped using the iCOGS array. After quality control of the phenotypes and genotypes, data were available on 15,252 BRCA1 mutation carriers, of whom 7,797 had been diagnosed with breast cancer, all of European ancestry. Analyses in BRCA1 mutation carriers assessed associations with breast cancer risk. (iii) Mammographic density information was available for 7,025 women from ten studies in BCAC and, in addition, 1,621 women from the Mayo Mammographic Health Study (MMHS). All were additionally participants in the MODE Consortium. Forty-six women were excluded because of missing BMI information, leaving 8,600 women with mammographic density information, relevant covariates and iCOGS genotyping (2,955 breast cancer cases and 5,645 controls). Study details are given in Supplementary Table 14 and in Lindstrom et al.26. Mammographic density measurements were performed on digitized analog mammographic films using ‘Cumulus’ software. This applies a thresholding technique to measure the total area of the breast and the absolute dense area, from which the absolute non-dense area and percent dense area are derived. Dense areas and non-dense areas were converted to cm² according to the pixel size used in the digitization. Readers blinded to genotype, case status and risk factor data conducted all measures. For cases, mammograms before the diagnosis of breast cancer were used or, where this was not possible, measures from the contralateral breast were used.

SNP selection, genotyping and imputation

We first defined a mapping interval of ~1 Mb (chr. 6: 151,600,000–152,650,000; NCBI Build 37 assembly). We catalogued 2,821 variants with a MAF >2% using the 1000 Genomes Project (March 2010 Pilot version 60 CEU project data); of these variants, we selected 277 SNPs correlated with the 3 previously reported associated SNPs (rs2046210 (ref. 1), rs3757318 (ref. 2) and rs3020314 (ref. 40)) at $r^2$ >0.1, plus a set of 698 SNPs designed to tag all remaining SNPs with $r^2$ >0.9. Of the SNPs, 902 that passed quality control were included in this analysis. After completion of iCOGS genotyping, this initial set was supplemented with a further 23 SNPs selected from the October 2010 (Build 37) release of the 1000 Genomes Project, to improve coverage. These SNPs were genotyped in two large BCAC studies (CCHS and SEARCH) comprising 12,273 cases and controls, using a Fluidigm array according to the manufacturer’s instructions. Using the above data, results for all the additional known common variants (MAF >0.02 in Europeans) on the January 2012 release of the 1000 Genomes Project were imputed using IMPUTE version 2.0. Quality control and imputation steps were carried out separately in the different consortia, leading to slight differences in the numbers of SNPs with available data. In addition to the 902 successfully genotyped SNPs, genotypes at 2,972 SNPs were imputed in BCAC and 2,907 SNPs were imputed in CIMBA (imputation $r^2$ score >0.3 in each case). In total, 3,872 genotyped or imputed SNPs were available for the combined BCAC ER− and CIMBA BRCA1 mutation carrier meta-analysis.
Statistical analysis

Case-control analysis, logistic regression and retrospective cohort analyses—For the case-control analysis in BCAC, per-allele odds ratios and standard errors were estimated for each SNP using logistic regression, separately for subjects of European and Asian ancestry and for each tested phenotype. Principal components were included as covariates as previously described\textsuperscript{21}. The statistical significance of each SNP was derived using a Wald test. To evaluate evidence for multiple association signals, we performed conditional analyses in which the association for each SNP was reevaluated after including other associated SNPs in the model. SNPs with a $P$ value $<1 \times 10^{-4}$ and MAF $>2\%$ in the single-SNP analysis were included in this analysis\textsuperscript{21}. Haplotype-specific odds ratios and confidence limits were estimated using haplo.stats\textsuperscript{22}.

Associations between genotypes and breast cancer risk in $BRCA1$ mutation carriers in CIMBA were evaluated using a per-allele trend test with 1 degree of freedom ($P_{\text{trend}}$), based on modeling the retrospective likelihood of the observed genotypes conditional on breast cancer phenotypes\textsuperscript{41}. To allow for non-independence among related individuals, an adjusted test statistic was used that took into account the correlation in genotypes\textsuperscript{21}. Per-allele hazard ratio estimates were obtained by maximizing the retrospective likelihood. All analyses were stratified by country of residence.

Conditional analyses were performed to identify SNPs independently associated with each phenotype. To identify the most parsimonious model, all SNPs with a marginal $P$ value $<1 \times 10^{-4}$ were included in forward selection regression analyses with a threshold for inclusion of $P < 1 \times 10^{-4}$ and including terms for principal components and study. Similarly, forward selection Cox regression analysis was performed for $BRCA1$ mutation carriers, stratified by country of residence, using the same $P$-value thresholds. This approach provides valid significance tests of the associations, although the estimates quantifying the association can be biased\textsuperscript{41,42}. Parameter estimates for the most parsimonious model were obtained using the retrospective likelihood approach.

Within MODE, mammographic dense area, non-dense area and percent dense area were each square-root transformed to fit a normal distribution. For the ten MODE and BCAC studies, a linear regression assuming a multiplicative per-allele model adjusting for study, age at mammogram, BMI, menopausal status (pre or post) and the first six principal components was carried out for each trait and for each SNP. The MMHS participants were analyzed separately in the same way but without the principal-components covariates, and the results were combined with those from BCAC using a standard inverse variance–weighted fixed-effects meta-analysis.

Expression analysis

eQTL analyses were conducted in 57 normal breast samples from the GTEx Project\textsuperscript{43} and 135 adjacent normal breast samples from women of European origin in the METABRIC study\textsuperscript{44}. For the METABRIC analyses, matched gene expression (Illumina HT-12 v3 microarray) and germline SNP data from either genotyping (Affymetrix SNP 6.0) or imputation (1000 Genomes Project, March 2012 data using IMPUTE version 2.0) were
used. Correlations between the five signal-representative SNPs and expression levels of nearby genes (500 kb upstream and downstream of the SNPs) were assessed using a linear regression model in which an additive effect on expression level was assumed for each copy of the rare allele. Calculations were carried out using the eMAP library in R.

**Allele-specific expression analysis**

ASE analysis has been described previously\(^\text{11}\). Three SNPs for signal 1, two SNPs for signal 3 and a proxy SNP for signal 2 (\(r^2 = 0.85\)) were on Affymetrix SNP Array 6.0. TCGA genotype calls and corresponding confidence scores were retrieved using level 2 TCGA SNP array Birdseed data downloaded from the TCGA portal. Genotyping data with a confidence score of 0.1 were excluded. We selected 742 breast cancer samples with European ancestry. The corresponding RNA-seq BAM files and metadata are available from the Cancer Genomics Hub (CGHub). Marker SNPs, the exonic SNPs of the target genes, were extracted from dbSNP human Build 142 (collectively ~800 SNPs for *ESRI, RMND1, ARMT1* and *CCDC170*), and RNA-seq read counts on SNP sites for reference and alternative alleles were computed. Homozygote marker SNPs and those with low coverage (less than 15\(\times\)) were excluded. Major allele fraction (\(\mu\)) representing allelic imbalance for each marker SNP was computed, and an average of allelic imbalances for each gene was calculated for individual tumor samples. Marker SNPs with extreme \(\mu\) values (\(\mu > 0.75\)) were not included in the analysis. Level 3 SNP array data were downloaded from the TCGA portal, and GISTIC version 2.0.16 was used to identify copy number variations (CNVs) for each sample. Samples with low or high CNV levels, as presented in the gene-based GISTIC module report, were excluded from the analysis of the corresponding gene. For each risk SNP, allelic imbalance for the target transcripts was compared between heterozygote (AB) and homozygote (AA and BB) samples. For a given risk SNP and target gene, we used Levene’s test, a more robust test than the \(F\) test, for equality of variances when the risk SNP was not in linkage disequilibrium with any of the marker SNPs on that gene (\(r^2 < 0.5\)). Otherwise, a two-tailed \(t\) test was used for equality of means\(^{45}\).

**Estrogen receptor protein expression**

Normal breast samples derived from 150 postmenopausal donors (non-Hispanic, mean age of 62 years) and identified through the Susan G. Komen for the Cure Tissue Bank at the Indiana University Simon Cancer Center were used in this study\(^{46}\). DNA was extracted from blood cells at the Indiana CTSI Specimen Storage Facility using an AutogenFlex Star instrument (Autogen) and the FlexiGene AGF3000 blood kit for DNA extractions (Qiagen). SNP analysis was performed with 1 ng of DNA using TaqMan genotyping assays for rs2046210 (C\(_{12034236\_10}\)), rs3757322 (C\(_{27475059\_10}\)), rs9397437 (C\(_{11556300\_10}\)), rs851984 (C\(_{2496819\_10}\)), rs9918437 (C\(_{29496189\_10}\)) and rs2747652 (C\(_{2823750\_10}\)) from Life Technologies, following the manufacturer’s protocol. ER protein abundance was measured by immunohistochemical semi-quantitation using an antibody to ER\(\alpha\) (clone 6F11; 1:40 dilution; Leica Microsystems) and quantified with (i) an H score consisting of the sum of the percent of tumor cells staining, multiplied by an ordinal value corresponding to the intensity level (0, none; 1, weak; 2, moderate; 3, strong; Supplementary Fig. 2), and (ii) the percentage of positive cells. Correlations between the H scores and ER immunohistochemistry values were calculated using Spearman’s rank correlation analysis.
All $P$ values reported are two-sided, and values <0.05 were considered statistically significant.

**Cell lines**

Breast cancer cell lines MCF-7 (ER$^+\); American Type Culture Collection (ATCC) HTB22), T-47D (ER$^+\); ATCC HTB133) and BT-474 (ER$^+\); ATCC HTB20) were grown in RPMI medium with 10% FCS and antibiotics. Normal breast epithelial cell lines MCF10A (ATCC CRL 10317) and Bre-80 (provided as a gift from R. Reddel, Children’s Medical Research Institute, Sydney) were grown in DMEM/F12 medium with 5% horse serum (HS), 10 μg/ml insulin, 0.5 μg/ml hydrocortisone, 20 ng/ml epidermal growth factor, 100 ng/ml cholera toxin and antibiotics. Cell lines were maintained under standard conditions, routinely tested for mycoplasma and short tandem repeat (STR) profiled.

**Chromatin conformation capture**

3C libraries were generated using EcoRI, HindIII or BglII as described previously$^{15}$. 3C interactions were quantified by RT-PCR (qPCR) using primers designed within restriction fragments (Supplementary Table 15). qPCR was performed on a RotorGene 6000 instrument using MyTaq HS DNA polymerase (Bioline) with the addition of 5 mM Syto9, an annealing temperature of 66 °C and an extension time of 30 s. 3C analyses were performed in three independent 3C libraries from each cell line, with each experiment quantified in duplicate. BAC clones (RP11-108N8, RP11-713G5, RP11-450E24 and RP11-55K19) covering the 6q25 region were used to create artificial libraries of ligation products to normalize for PCR efficiency. Data were normalized to the signal from the BAC clone library and, between cell lines, by reference to a region within GAPDH. All qPCR products were electrophoresed on 2% agarose gels, gel purified and sequenced to verify the 3C product.

**Electromobility shift assays**

Gel shift assays were performed with ER$^+$ MCF-7 or ER$^-$ Bre-80 nuclear lysates and biotinylated oligonucleotide duplexes (Supplementary Table 16). Nuclear lysates were prepared using NE-PER nuclear and cytoplasmic extraction reagents (Thermo Fisher Scientific) according to the manufacturer’s instructions. Total protein concentrations in nuclear lysates were determined by Bradford’s method. Duplexes were prepared by combining sense and antisense oligonucleotides in NEBuffer2 (New England BioLabs) and heat annealing at 80 °C for 10 min followed by slow cooling to 25 °C for 1 h. Binding reactions were performed in binding buffer (10% glycerol, 20 mM HEPES (pH 7.4), 1 mM DTT, protease inhibitor cocktail (Roche), 0.75 μg poly(dI:dC) (Sigma-Aldrich)) with 7.5 μg of nuclear lysate. For competition assays, binding reactions were preincubated with 1 pmol of competitor duplex (Supplementary Table 17) at 25 °C for 10 min before the addition of 10 fmol of biotinylated oligonucleotide duplex and a further incubation at 25 °C for 15 min. Reactions were separated on 10% Tris-borate-EDTA (TBE) polyacrylamide gels (Bio-Rad) in TBE buffer at 160 V for 40 min. Duplex-bound complexes were transferred onto Zeta-Probe positively charged nylon membranes (Bio-Rad) by semidry transfer at 25 V for 20 min and then cross-linked onto the membranes under 254-nm ultraviolet light for 10 min. Membranes were processed with the LightShift Chemiluminescent EMSA kit (Thermo...

*Nat Genet. Author manuscript; available in PMC 2016 October 01.*
Fisher Scientific) according to the manufacturer’s instructions. Chemiluminescent signals were visualized with the C-DiGit blot scanner (LI-COR).

**Plasmid construction and reporter assays**

Promoter-driven luciferase reporter constructs were generated by the insertion of PCR-amplified fragments containing *ESR1, ARMT1, RMND1* and *CCDC170* promoters into the KpnI and MluI sites of pGL3-Basic. To assist in cloning, AgeI and SbfI sites were inserted into the BamHI and SalI sites downstream of the luciferase gene. A 1,496-bp signal 1 PRE fragment, a 997-bp signal 2 PRE fragment, a 1,566-bp signal 3 PRE fragment, a 1,463-bp signal 4 PRE fragment and a 1,349-bp signal 5 PRE fragment were generated by PCR or gBlocks (Integrated DNA Technologies) and cloned into the AgeI and SbfI sites of the modified pGL3-Promoter constructs. The minor alleles of individual SNPs were introduced into the PRE sequences by overlap extension PCR or gBlocks. Sequencing of all constructs confirmed variant incorporation (AGRF). ER⁺ MCF-7 and BT-474 or ER⁻ Bre-80 cells were transfected with equimolar amounts of luciferase reporter plasmids and 50 ng of pRL-TK transfection control plasmid with Lipofectamine 3000. The total amount of transfected DNA was kept constant at 600 ng for each construct by the addition of pUC19 as a carrier plasmid. Luciferase activity was measured 24 h after transfection by the Dual-Glo Luciferase Assay System. To correct for any differences in transfection efficiency or cell lysate preparation, firefly luciferase activity was normalized to *Renilla* luciferase activity, and the activity of each construct was measured relative to the promoter-only construct, which had a defined activity of 1. Statistical significance was tested by log transforming the data and performing two-way ANOVA followed by Dunnett’s multiple-comparisons test in GraphPad Prism.

**Chromatin immunoprecipitation**

ER⁺ MCF-7 and BT-474 breast cancer cells were cross-linked with 1% formaldehyde at 37 °C for 10 min, rinsed once with ice-cold PBS containing 5% BSA and once with PBS, and collected in PBS containing 1x protease inhibitor cocktail (Roche). The cells were centrifuged for 2 min at 900g. Cell pellets were resuspended in 0.35 ml of lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1, 1x pro-tease inhibitor cocktail) and sonicated three times for 15 s each with a 70% duty cycle (Branson SLPt) followed by centrifugation at 15,000g for 15 min. Supernatants were collected and diluted in dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl, pH 8.1). Two micrograms of antibody was prebound for 6 h to Protein G Dynabeads (Life Technologies) and then added to the diluted chromatin for overnight immunoprecipitation. The magnetic bead–chromatin complexes were collected and washed six times in RIPA buffer (50 mM HEPES, pH 7.6, 1 mM EDTA, 0.7% sodium deoxy-cholate, 1% NP-40, 0.5 M LiCl) and then twice with TE buffer. To reverse the cross-linking, the magnetic bead complexes were incubated overnight at 65 °C in elution buffer (1% SDS, 0.1 M NaHCO₃). DNA fragments were purified using a QIAquick Spin kit (Qiagen). For qPCR, 2.0 μl from a 100-μl immu-noprecipitated chromatin extraction was subjected to 40 cycles of amplification. All PCR products were sequenced by Sanger sequencing (AGRF). The antibodies used were to CTCF (C-20; sc-15914) and GATA3 (HG3-31; sc268) or control IgG (sc-2027) (all from Santa Cruz Biotechnology). ChIP primers are listed in Supplementary Table 18.
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Affiliations

1Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK 2Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
3Cancer Division, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia 4Breast Cancer Research, Breakthrough Breast Cancer Research Centre, London, UK 5Academic Biochemistry, Royal Marsden Hospital, London, UK 6Department of Genetics, Stanford University School of Medicine, Stanford, California, USA 7Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA 8Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands 9Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia 10Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia 11Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, UK 12Institute of Population Health, University of Manchester, Manchester, UK 13Department of Gynecology and Obstetrics, University of Erlangen, Friedrich
Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-Nuremberg Metropolitan Region, Erlangen, Germany
Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA
Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
Research Oncology, Division of Cancer Studies, King’s College London, Guy’s Hospital, London, UK
Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
Oxford Biomedical Research Centre, University of Oxford, Oxford, UK
Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany
Molecular Epidemiology Group, German Cancer Research Center, Heidelberg, Germany
National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany
Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany
Environmental Epidemiology of Cancer, Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France
University Paris–Sud, Villejuif, France
Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark
Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark
Human Cancer Genetics Program, Spanish National Cancer Centre (CNIO), Madrid, Spain
Servicio de Cirugía General y Especialidades, Hospital Monte Naranco, Oviedo, Spain
Department of Epidemiology, University of California–Irvine, Irvine, California, USA
Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA
Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany
German Cancer Consortium, German Cancer Research Center, Heidelberg, Germany
Department of Gynaecology and Obstetrics, Technical University of Munich, Munich, Germany
Division of Molecular Gynecology, Department of Gynaecology and Obstetrics, University Hospital of Cologne, Cologne, Germany
Centre of Familial Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany
Center for Integrated Oncology, University Hospital, Cologne, Germany
Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany
University of Tübingen, Tübingen, Germany
Molecular Genetics of Breast Cancer, German Cancer Research Center, Heidelberg, Germany
Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland
Department of Oncology, Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland
Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan
Division of Molecular Medicine, Aichi Cancer Center Research Institute, Nagoya, Japan
Radiation Oncology Research Unit, Hannover Medical School, Hannover, Germany
Gynaecology Research Unit, Hannover Medical School, Hannover, Germany
Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
Department of Oncology-Pathology, Karolinska University
Hospital, Stockholm, Sweden 50-Cancer Center, Kuopio University Hospital, Kuopio, Finland 51-Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland 52-Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland 53-Vesalius Research Center, Leuven, Belgium 54-Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium 55-Multidisciplinary Breast Center, Department of General Medical Oncology, University Hospitals Leuven, Leuven, Belgium 56-Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany 57-University Cancer Center Hamburg (UCC), University Medical Center Hamburg-Eppendorf, Hamburg, Germany 58-FOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy 59-Unit of Molecular Basis of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy 60-Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA 61-Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia 62-Department of Medicine, McGill University, Montreal, Quebec, Canada 63-Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montreal, Quebec, Canada 64-Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Malaysia 65-Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya Medical Centre, Kuala Lumpur, Malaysia 66-Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Oslo, Norway 67-Institute of Clinical Medicine, University of Oslo, Oslo, Norway 68-Department of Clinical Molecular Biology, Oslo University Hospital, University of Oslo, Oslo, Norway 69-Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA 70-Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, NordLab Oulu University Hospital, Oulu, Finland 71-Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, Oulu, Finland 72-Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada 73-Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada 74-Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada 75-Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada 76-Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands 77-Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands 78-Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, the Netherlands 79-Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA 80-Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden 81-Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands 82-Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China 83-Sheffield Cancer Research, Department of Oncology, University of Sheffield, Sheffield, UK
Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK
International Epidemiology Institute, Rockville, Maryland, USA
Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea
Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea
Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea
Department of Haematology-Oncology, National University Health System, Singapore
Cancer Science Institute of Singapore, National University of Singapore, Singapore
Saw Swee Hock School of Public Health, National University of Singapore, Singapore
Department of Surgery, National University Health System, Singapore
Institute of Human Genetics, Pontificia Universidad Javerianar, Bogota, Colombia
Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
International Agency for Research on Cancer, Lyon, France
National Cancer Institute, Bangkok, Thailand
Roswell Park Cancer Institute, Buffalo, New York, USA
Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA
School of Public Health, China Medical University, Taichung, Taiwan
Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan
Division of Cancer Studies, Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, UK
Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK
Division of Breast Cancer Research, Institute of Cancer Research, London, UK
Centre for Epidemiology and Biostatistics, University of Melbourne, Melbourne, Victoria, Australia
Department of Epidemiology, Cancer Prevention Institute of California, Fremont, California, USA
Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA
Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, Pennsylvania, USA
Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah, USA
Department of Medicine, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah, USA
State Research Institute Centre for Innovative Medicine, Vilnius, Lithuania
Latvian Biomedical Research and Study Centre, Riga, Latvia
Department of Medical Oncology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
Department of Genetics, University of Pretoria, Pretoria, South Africa
Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, California, USA
Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
Human Genetics Group, Spanish National Cancer Centre (CNIO), Madrid, Spain
Biomedical Network on Rare Diseases (CIBERER), Madrid, Spain
Human Genotyping (CEGEN) Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain
City of Hope Clinical Cancer Genomics Community Research Network, Duarte, California, USA
Clinical Cancer Genetics, City of Hope, Duarte, California, USA
Division of Cancer Prevention and
Genetics, Istituto Europeo di Oncologia, Milan, Italy
Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione Istituto di Ricovery e Cura a Carattere Scientifico, Istituto Nazionale Tumori, Milan, Italy
Unit of Medical Genetics, Department of Biomedical, Experimental and Clinical Sciences, University of Florence, Florence, Italy
Department of Molecular Medicine, University La Sapienza, Rome, Italy
Molecular Diagnostics Laboratory, INRASTES (Institute of Nuclear and Radiological Sciences and Technology), National Centre for Scientific Research 'Demokritos', Aghia Paraskevi Attikis, Athens, Greece
Cancer Risk and Prevention Clinic, Dana-Farber Cancer Institute, Boston, Massachusetts, USA
Department of Basic Sciences, Shaukat Khanum Memorial Cancer Hospital and Research Centre, Lahore, Pakistan
Clinical Genetics, Guy's and St. Thomas' National Health Service (NHS) Foundation Trust, London, UK
Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA
Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Campus Kiel, Christian Albrechts University Kiel, Kiel, Germany
University Düsseldorf, Dusseldorf, Germany
Centre of Familial Breast and Ovarian Cancer, Department of Gynaecology and Obstetrics and Centre for Integrated Oncology, Center for Molecular Medicine Cologne, University Hospital of Cologne, Cologne, Germany
Institute of Human Genetics, Münster, Germany
INSERM U1052, CNRS UMR 5286, Université Lyon, Centre de Recherche en Cancérologie de Lyon, Lyon, France
Institut Curie, Department of Tumour Biology, Paris, France
Université Paris Descartes, Sorbonne Paris Cité, Paris, France
Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon–Centre Léon Bérard, Lyon, France
Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA
Center for Medical Genetics, Ghent University, Ghent, Belgium
Molecular Oncology Laboratory, Hospital Clínico San Carlos, IdISSC (El Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain
Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland
Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands
Department of Medical Oncology, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, the Netherlands
Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, the Netherlands
Hereditary Breast Cancer Family Registry, Cancer Genetics Center, Hong Kong
Sanatorium and Hospital, Hong Kong
Department of Surgery, University of Hong Kong
Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary
Oncogenetics Laboratory, Vall d’Hebron Institute of Oncology (VHIO), Vall d’Hebron University Hospital, Barcelona, Spain
Genetic Counseling Unit, Hereditary Cancer Program, IDIBGI (Institut d’Investigació Biomèdica de Girona), Catalan Institute of Oncology, Girona, Spain
Breast Cancer and Systems Biology Unit, IDIBELL (Bellvitge Biomedical Research Institute), Catalan Institute of Oncology, Barcelona, Spain
Department of Pathology, Landspitali University Hospital and Biomedical Centre (BMC), Faculty of
Medicine, University of Iceland, Reykjavik, Iceland 154Medical Genetic Division, Centre Hospitalier Universitaire de Québec and Laval University, Quebec City, Quebec, Canada 155Centre Hospitalier Universitaire de Québec and Laval University, Quebec City, Quebec, Canada 156Immunology and Molecular Oncology Unit, Istituto Oncologico Veneto (IOV), IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico), Padua, Italy 157Department of Genetics, Portuguese Oncology Institute, Porto, Portugal 158Biomedical Sciences Institute (ICBAS), Porto University, Porto, Portugal 159Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA 160Program in Cancer Genetics, McGill University, Montreal, Quebec, Canada 161Masaryk Memorial Cancer Institute and Medical Faculty, Masaryk University, Brno, Czech Republic 162Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA 163Department of Obstetrics and Gynecology, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria 164Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, Florida, USA 165Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Rockville, Maryland, USA 166Department of Community Medicine and Epidemiology, Carmel Medical Center and B. Rappaport Faculty of Medicine, Haifa, Israel 167Clalit National Israeli Cancer Control Center, Haifa, Israel 168N.N. Petrov Institute of Oncology, St. Petersburg, Russia 169Center for Medical Genetics, NorthShore University Health System, Evanston, Illinois, USA 170Division of Cancer Medicine, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia 171NRG Oncology, Statistics and Data Management Center, Roswell Park Cancer Institute, Buffalo, New York, USA 172Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada 173Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada 174Department of Clinical Genetics, Vejle Hospital, Vejle, Denmark 175Department of Clinical Genetics, Odense University Hospital, Odense, Denmark 176Section of Genetic Oncology, Department of Laboratory Medicine, University of Pisa and University Hospital of Pisa, Pisa, Italy 177University Malaya Cancer Research Institute, Faculty of Medicine, University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia 178Susanne Levy Gertner Oncogenetics Unit, Sheba Medical Center, Tel-Hashomer, Israel 179Department of Oncology, Lund University, Lund, Sweden 180Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden 181Department of Clinical Genetics, Lund University Hospital, Lund, Sweden 182Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden 183Center for Clinical Cancer Genetics and Global Health, University of Chicago Medical Center, Chicago, Illinois, USA 184Division of Cancer Prevention and Control Research, Jonsson Comprehensive Cancer Center, University of California Los Angeles School of Medicine and School of Public Health, Los Angeles, California, USA 185Department of Medicine and Genetics, University of California, San Francisco, San Francisco, California, USA 186Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA 187University of Texas MD Anderson
Cancer Center, Houston, Texas, USA 188Familial Cancer Centre, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia 188Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia 190Women’s Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA 191University of Hawaii Cancer Center, Honolulu, Hawaii, USA 192Department of Obstetrics, Gynaecology and Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada 193Centre for Genetic Origins of Health and Disease, University of Western Australia, Perth, Western Australia, Australia 194Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts, USA 195Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA 196Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, Massachusetts, USA 197Clinical Gerontology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Acknowledgments

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. This study would not have been possible without the contributions of the following: A. Berchuck (OCAC), R.A. Eeles, A.A. Al Olama, Z. Kote-Jarai and S. Benlloch (PRACTICAL), C. Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, the staff of the CNIO genotyping unit, D.C. Tessier, F. Bacot, D. Vincent, S. LaBoissière, F. Robidoux and the staff of the McGill University and Génomique Québec Innovation Centre, S.F. Nielsen, B.G. Nordestgaard and the staff of the Copenhagen DNA laboratory, and J.M. Cunningham, S.A. Windebank, C.A. Hilker, J. Meyer and the staff of the Mayo Clinic Genotyping Core Facility. Normal human tissues from the Susan G. Komen for the Cure Tissue Bank at the Indiana University Simon Cancer Center (Indianapolis) were used in this study. We thank the contributors, including Indiana University who collected samples used in this study, as well as the donors and their families, whose help and participation made this work possible. We also acknowledge National Institute for Health Research (NIHR) support to the Royal Marsden Biomedical Research Centre. Funding for the iCOGS infrastructure came from the European Community’s Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692 and C8197/A16565), the US National Institutes of Health (NIH; CA128978, CA192393, CA16167, CA176785 and an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201)) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112, the GAME-ON initiative), the US Department of Defense (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, the Komen Foundation for the Cure, the Breast Cancer Research Foundation and the Ovarian Cancer Research Fund. Full acknowledgments are given in the Supplementary Note.

References


Figure 1.
Association results for all SNPs with six phenotypes. (a–f) The phenotypes analyzed include risk of ER$^+$ breast cancer in BCAC (a), risk of ER$^-$ breast cancer in BCAC (b), risk of triple-negative breast cancer, derived from the CIMBA meta-analysis of BRCA1 mutation carriers with ER$^-$ tumors (c), risk of HER2$^+$ breast cancer in BCAC (d), mammographic dense area in MODE (e) and tumor grade after adjustment for ER status in BCAC (f). $P$ values for each SNP (from unconditional logistic regression) are shown plotted as the negative log-transformed $P$ value against relative position across the locus. A schematic of the gene structures is shown above a and d. The physical positions of signals 1–5 are shown as colored, numbered stripes. Dotted horizontal lines indicate the genome-wide significance level.
Figure 2.
ER expression and allelic imbalance correlate with signal 1 SNPs. (a) Negative correlation between the signal 1 SNP rs2046210 and ER protein expression. Black dots represent ER expression from individual samples measured by immunohistochemistry (H score). Horizontal lines represent the mean H score for each genotype. The $P$ value was calculated using a Spearman rank correlation test. (b) Box plots of ESR1 gene expression ($\log_2$ transformed) in breast tumor and adjacent normal samples. Boxes extend from the 25th to the 75th percentile, horizontal bars represent the median, whiskers indicate the full range of ESR1 expression and outliers are represented as circles. (c) Allelic imbalance in ESR1 expression by genotypic status at breast cancer risk variants. Data are classified according to the genotypes at risk SNPs (heterozygous versus homozygous). Black dots represent the average major allele fraction of the marker SNPs across ESR1 for an individual from TCGA with breast cancer. Red lines and whiskers correspond to means $\pm$ 1 s.d. For rs7740686 (signal 1) and rs9397437 (signal 2), Levene’s test (equality of variances) was used to calculate the $P$ values; for rs851985 (signal 3), a two-tailed $t$ test (equality of means) was used to calculate the $P$ value.
Figure 3.
Chromatin interactions across the 6q25.1 risk region. (a) Signals 1–5 are numbered and shown as colored stripes. RMND1, ARMT1, CCDC170 and ESR1 gene structures are depicted with exons (vertical bars) joined by introns (lines). Gene-enhancer predictions from PreSTIGE\textsuperscript{12}, ChIP-seq binding profiles for H3K27ac\textsuperscript{13} and Encyclopedia of DNA Elements (ENCODE) RNA polymerase II ChIA-PET interactions in MCF-7 cells are shown. (b–d) 3C anchor points (3C baits) and interrogated sequences (3C regions) are depicted as black boxes and gray shading, respectively. 3C interaction profiles in ER\textsuperscript{+} MCF-7 and ER\textsuperscript{−} Bre-80 breast cell lines are shown for signals 1 and 2 (b), signals 3 and 4 (c) and signal 5 (d). 3C libraries were generated with EcoRI, with the anchor point set at the ESR1, RMND1-ARMT1 or
*CCDC170* promoter region. Graphs present the results from three biological replicates; error bars, s.d.
Figure 4. Risk alleles reduce *ESR1* and *RMND1* promoter activity. Luciferase reporter assays were performed following transient transfection of ER*+* MCF-7 breast cancer cells. PREs containing the major SNP alleles were cloned downstream of target gene promoter-driven luciferase constructs (prom) for the creation of reference (Ref-PRE) constructs. Minor SNP alleles were engineered into the constructs and are designated by the rsID of the corresponding SNP. “Haplotype” denotes a construct that contains the minor alleles of both candidate SNPs within signal 1 or 3. Error bars, 95% confidence intervals from three independent experiments. *P* values were determined by two-way ANOVA followed by Dunnett’s multiple-comparisons test: **$P < 0.01$, ***$P < 0.001$. 
Figure 5.
GATA3 and CTCF binding in vivo. (a) ChIP and quantitative PCR (qPCR) assays using antibody against GATA3 or CTCF in ER+ BT-474 breast cancer cells. A region within the second intron of ESR1 served as a negative control (NC). Normal rabbit IgG was used as a non-specific antibody control. Graphs present the results of two biological replicates; error bars, s.d. (b) 3C followed by sequencing for the signal 4 PRE containing rs1361024 in heterozygous ER+ MCF-7 breast cancer cells shows allele-specific chromatin looping. The chromatograms represent one of three independent 3C libraries generated and sequenced.
### Table 1

The associations of each signal-representative SNP with tumor risk and mammographic density in the three contributing consortia

<table>
<thead>
<tr>
<th>Signal</th>
<th>Representative SNP</th>
<th>Position</th>
<th>Alleles</th>
<th>Freq.</th>
<th>OR (95% CI)</th>
<th>cOR (95% CI)</th>
<th>P trend</th>
<th>cOR (95% CI)</th>
<th>OR (95% CI)</th>
<th>cOR (95% CI)</th>
<th>P cond</th>
<th>HR (95% CI)</th>
<th>cHR (95% CI)</th>
<th>P value</th>
<th>cHR (95% CI)</th>
<th>P cond</th>
<th>Meta-analysis P value</th>
<th>P cond</th>
<th>( \beta ) (95% CI)</th>
<th>c( \beta ) (95% CI)</th>
<th>P trend</th>
<th>c( \beta ) (95% CI)</th>
<th>P cond</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs3757322</td>
<td>151,942,194</td>
<td>GT</td>
<td>0.33</td>
<td>1.17 (1.12–1.21)</td>
<td>1.00 x 10^{-14}</td>
<td>1.07 (1.04–1.09)</td>
<td>1.10 x 10^{-10}</td>
<td>1.15 (1.10–1.20)</td>
<td>3.78 x 10^{-3}</td>
<td>2.50 x 10^{-23}</td>
<td>0.12 (0.07, 0.17)</td>
<td>1.82 x 10^{-6}</td>
<td>1.14 (1.10–1.19)</td>
<td>1.51 x 10^{-9}</td>
<td>1.06 (1.04–1.09)</td>
<td>1.02 x 10^{-9}</td>
<td>1.10 (1.06–1.14)</td>
<td>3.79 x 10^{-7}</td>
<td>7.59 x 10^{-15}</td>
<td>0.07 (0.01, 0.12)</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>rs9397437</td>
<td>151,952,332</td>
<td>AG</td>
<td>0.07</td>
<td>1.28 (1.19–1.37)</td>
<td>5.29 x 10^{-12}</td>
<td>1.15 (1.10–1.20)</td>
<td>1.26 x 10^{-9}</td>
<td>1.24 (1.15–1.33)</td>
<td>3.98 x 10^{-8}</td>
<td>1.24 x 10^{-19}</td>
<td>0.27 (0.18, 0.36)</td>
<td>2.36 x 10^{-9}</td>
<td>1.18 (1.11–1.26)</td>
<td>1.20 x 10^{-5}</td>
<td>1.12 (1.07–1.17)</td>
<td>3.56 x 10^{-6}</td>
<td>1.12 (1.05–1.19)</td>
<td>3.60 x 10^{-4}</td>
<td>3.29 x 10^{-8}</td>
<td>0.22 (0.12, 0.32)</td>
<td>1.66 x 10^{-5}</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>rs851984</td>
<td>152,023,191</td>
<td>AG</td>
<td>0.41</td>
<td>1.04 (1.01–1.08)</td>
<td>0.024</td>
<td>1.06 (1.03–1.08)</td>
<td>1.97 x 10^{-6}</td>
<td>1.05 (1.01–1.10)</td>
<td>0.015</td>
<td>9.14 x 10^{-4}</td>
<td>-0.03 (−0.07, 0.02)</td>
<td>0.29</td>
<td>NA</td>
<td>1.07 (1.05–1.10)</td>
<td>1.09 x 10^{-9}</td>
<td>1.07 (1.03–1.10)</td>
<td>3.60 x 10^{-4}</td>
<td>3.12 x 10^{-5}</td>
<td>0.01 (−0.04, 0.06)</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>rs9918437</td>
<td>152,072,718</td>
<td>TG</td>
<td>0.07</td>
<td>1.18 (1.11–1.27)</td>
<td>6.20 x 10^{-7}</td>
<td>1.08 (1.04–1.13)</td>
<td>1.04 x 10^{-9}</td>
<td>1.17 (1.08–1.26)</td>
<td>1.30 x 10^{-4}</td>
<td>1.48 x 10^{-10}</td>
<td>0.03 (−0.05, 0.12)</td>
<td>0.45</td>
<td>1.13 (1.06–1.20)</td>
<td>4.46 x 10^{-4}</td>
<td>NA</td>
<td>1.10 (1.04–1.17)</td>
<td>0.0015</td>
<td>2.61 x 10^{-6}</td>
<td>0.03 (−0.06, 0.12)</td>
<td>0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>rs2747652</td>
<td>152,437,016</td>
<td>CT</td>
<td>0.54</td>
<td>1.12 (1.08–1.16)</td>
<td>1.83 x 10^{-9}</td>
<td>1.05 (1.03–1.08)</td>
<td>9.49 x 10^{-6}</td>
<td>1.00 (0.96–1.04)</td>
<td>0.95</td>
<td>1.44 x 10^{-5}</td>
<td>-0.02 (−0.07, 0.03)</td>
<td>0.39</td>
<td>1.12 (1.08–1.16)</td>
<td>2.32 x 10^{-9}</td>
<td>1.05 (1.03–1.08)</td>
<td>6.60 x 10^{-6}</td>
<td>1.00 (0.97–1.04)</td>
<td>0.86</td>
<td>5.97 x 10^{-5}</td>
<td>-0.02 (−0.07, 0.03)</td>
<td>0.45</td>
<td></td>
</tr>
</tbody>
</table>

For each signal-representative SNP (the best associated genotyped SNP), odds ratios for the minor/major allele and conditional odds ratios (OR and cOR) together with 95% confidence intervals (CIs), hazard ratios (HR and cHR), \( \beta \) coefficients (\( \beta \) and c\( \beta \) and \( P \) values (\( P \) cond)) are from models including the other four signal-representative SNPs. Representative conditional odds ratios and confidence
intervals could not be generated from the meta-analysis. NA indicates that the SNP was not included in conditional analysis because its individual effect was not significant at $P > 1 \times 10^{-4}$. Freq., frequency.

Mammographic dense area was square-root transformed and adjusted for age, BMI, menopausal status, study and relevant principal components.
The association of each signal-representative SNP with the main tumor subtype combinations and tumor grade

<table>
<thead>
<tr>
<th>Signal Representative SNP</th>
<th>n cases</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER/PR+/PR-HER2*</td>
<td>10,834</td>
<td>1.07 (1.03–1.11)</td>
<td>3.93 × 10^{-4}</td>
<td>1.14 (1.07–1.21)</td>
<td>9.54 × 10^{-5}</td>
<td>1.03 (0.92–1.06)</td>
<td>1.40 × 10^{-1}</td>
<td>1.10 (1.03–1.16)</td>
<td>4.16 × 10^{-3}</td>
<td>1.04 (1.00–1.07)</td>
<td>3.67 × 10^{-2}</td>
</tr>
<tr>
<td>ER/PR+/PR-HER2*</td>
<td>1.616</td>
<td>1.10 (1.02–1.19)</td>
<td>1.68 × 10^{-2}</td>
<td>1.25 (1.09–1.43)</td>
<td>1.05 × 10^{-3}</td>
<td>1.05 (0.98–1.13)</td>
<td>1.88 × 10^{-1}</td>
<td>1.05 (0.92–1.21)</td>
<td>4.75 × 10^{-1}</td>
<td>1.07 (0.99–1.15)</td>
<td>7.22 × 10^{-2}</td>
</tr>
<tr>
<td>Case-only P</td>
<td>6.60 × 10^{-1}</td>
<td>1.80 × 10^{-1}</td>
<td>3.90 × 10^{-1}</td>
<td>8.40 × 10^{-1}</td>
<td>4.00 × 10^{-1}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td>5,331</td>
<td>1.05 (1.00–1.10)</td>
<td>4.04 × 10^{-2}</td>
<td>1.04 (0.96–1.14)</td>
<td>3.17 × 10^{-1}</td>
<td>1.00 (0.96–1.05)</td>
<td>8.79 × 10^{-1}</td>
<td>1.07 (0.99–1.16)</td>
<td>7.54 × 10^{-2}</td>
<td>0.98 (0.95–1.03)</td>
<td>5.30 × 10^{-1}</td>
</tr>
<tr>
<td>Grade 2</td>
<td>11,498</td>
<td>1.08 (1.04–1.11)</td>
<td>8.77 × 10^{-6}</td>
<td>1.16 (1.09–1.23)</td>
<td>1.48 × 10^{-6}</td>
<td>1.05 (1.02–1.08)</td>
<td>3.51 × 10^{-3}</td>
<td>1.08 (1.02–1.14)</td>
<td>6.57 × 10^{-3}</td>
<td>1.06 (1.03–1.10)</td>
<td>7.91 × 10^{-5}</td>
</tr>
<tr>
<td>Grade 3</td>
<td>4,701</td>
<td>1.06 (1.01–1.11)</td>
<td>1.37 × 10^{-2}</td>
<td>1.17 (1.08–1.28)</td>
<td>2.22 × 10^{-4}</td>
<td>1.11 (1.06–1.16)</td>
<td>6.30 × 10^{-6}</td>
<td>1.16 (1.07–1.26)</td>
<td>2.17 × 10^{-4}</td>
<td>1.21 (1.08–1.17)</td>
<td>6.11 × 10^{-7}</td>
</tr>
<tr>
<td>Case-only P</td>
<td>9.20 × 10^{-1}</td>
<td>2.00 × 10^{-2}</td>
<td>2.60 × 10^{-4}</td>
<td>6.00 × 10^{-2}</td>
<td>7.97 × 10^{-6}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER/PR+HER2* (TN)</td>
<td>2,840</td>
<td>1.20 (1.12–1.28)</td>
<td>7.17 × 10^{-8}</td>
<td>1.25 (1.11–1.40)</td>
<td>1.50 × 10^{-4}</td>
<td>1.05 (0.98–1.12)</td>
<td>1.40 × 10^{-1}</td>
<td>1.17 (1.04–1.32)</td>
<td>7.00 × 10^{-3}</td>
<td>1.08 (1.01–1.15)</td>
<td>1.65 × 10^{-2}</td>
</tr>
<tr>
<td>ER/PR+HER2*</td>
<td>858</td>
<td>1.19 (1.07–1.32)</td>
<td>8.80 × 10^{-4}</td>
<td>1.25 (1.04–1.5)</td>
<td>1.55 × 10^{-2}</td>
<td>1.00 (0.91–1.11)</td>
<td>9.40 × 10^{-1}</td>
<td>1.18 (0.99–1.40)</td>
<td>6.80 × 10^{-2}</td>
<td>1.24 (1.12–1.37)</td>
<td>2.41 × 10^{-4}</td>
</tr>
<tr>
<td>Case-only P</td>
<td>7.80 × 10^{-1}</td>
<td>4.20 × 10^{-1}</td>
<td>1.40 × 10^{-1}</td>
<td>9.20 × 10^{-1}</td>
<td>2.08 × 10^{-2}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER-PR+HER2*</td>
<td>268</td>
<td>1.17 (0.97–1.40)</td>
<td>9.00 × 10^{-2}</td>
<td>1.14 (0.83–1.58)</td>
<td>4.10 × 10^{-1}</td>
<td>1.30 (1.10–1.55)</td>
<td>2.50 × 10^{-3}</td>
<td>1.14 (0.82–1.56)</td>
<td>4.40 × 10^{-3}</td>
<td>1.10 (0.92–1.31)</td>
<td>2.90 × 10^{-1}</td>
</tr>
<tr>
<td>Case only vs. TN</td>
<td>8.00 × 10^{-1}</td>
<td>7.80 × 10^{-1}</td>
<td>3.00 × 10^{-2}</td>
<td>6.50 × 10^{-1}</td>
<td>8.30 × 10^{-1}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case only vs. ER-PR-HER2*</td>
<td>6.40 × 10^{-1}</td>
<td>6.10 × 10^{-1}</td>
<td>3.00 × 10^{-2}</td>
<td>1.60 × 10^{-1}</td>
<td>3.70 × 10^{-1}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case only vs. ER-PR+HER2*</td>
<td>7.90 × 10^{-1}</td>
<td>7.60 × 10^{-1}</td>
<td>1.20 × 10^{-1}</td>
<td>9.90 × 10^{-1}</td>
<td>3.80 × 10^{-1}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td>218</td>
<td>1.23 (1.00–1.5)</td>
<td>4.40 × 10^{-2}</td>
<td>1.35 (0.96–1.91)</td>
<td>8.60 × 10^{-2}</td>
<td>0.87 (0.71–1.07)</td>
<td>1.60 × 10^{-1}</td>
<td>0.87 (0.6–1.26)</td>
<td>4.70 × 10^{-1}</td>
<td>1.01 (0.84–1.23)</td>
<td>9.00 × 10^{-1}</td>
</tr>
<tr>
<td>Grade 2</td>
<td>1,204</td>
<td>1.14 (1.05–1.24)</td>
<td>2.88 × 10^{-3}</td>
<td>1.19 (1.02–1.39)</td>
<td>2.63 × 10^{-2}</td>
<td>1.09 (0.99–1.18)</td>
<td>5.30 × 10^{-2}</td>
<td>1.26 (1.09–1.45)</td>
<td>1.79 × 10^{-3}</td>
<td>1.12 (1.03–1.22)</td>
<td>5.93 × 10^{-3}</td>
</tr>
<tr>
<td>Grade 3</td>
<td>3,463</td>
<td>1.20 (1.13–1.26)</td>
<td>6.10 × 10^{-31}</td>
<td>1.30 (1.19–1.43)</td>
<td>1.88 × 10^{-8}</td>
<td>1.05 (0.995–1.10)</td>
<td>7.49 × 10^{-2}</td>
<td>1.19 (1.09–1.31)</td>
<td>1.24 × 10^{-4}</td>
<td>1.12 (1.05–1.17)</td>
<td>4.36 × 10^{-5}</td>
</tr>
<tr>
<td>Grade polytomous adjusted for ER, constrained</td>
<td>9.18 × 10^{-1}</td>
<td>6.42 × 10^{-3}</td>
<td>5.43 × 10^{-4}</td>
<td>2.96 × 10^{-3}</td>
<td>1.82 × 10^{-3}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal Representative SNP</td>
<td>n cases</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3757322</td>
<td>OR (95% CI)</td>
<td>P</td>
<td>OR (95% CI)</td>
<td>P</td>
<td>OR (95% CI)</td>
<td>P</td>
<td>OR (95% CI)</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtypes with strongest association</td>
<td>ER−</td>
<td>High grade</td>
<td>High grade</td>
<td>ER−</td>
<td>ER+HER2+ and high grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bold text highlights the most significant phenotype associations within each signal. IHC, immunohistochemistry; TN, triple negative.
Table 3

Remaining candidate causal variants within each independent signal after likelihood-ratio testing based on exclusion phenotype

| Signal | Representative SNP | Exclusion phenotype | Lead SNP | Position | A1/A2 | Freq. | Conditional OR (95\% CI) | P<sub>trend</sub> overall breast cancer risk in BCAC | Unexcluded candidates | Chromosome position | P<sub>cond</sub> with lead SNP | P<sub>trend</sub> overall breast cancer risk in BCAC |
|--------|-------------------|-------------------|---------|----------|------|------|--------------------------|-----------------------|------------------------|----------------------|-------------------|------------------|------------------------|
| 1      | rs2046210         | Meta-analysis (BCAC ER<sup>-</sup> CIMBA) | rs2046210 | 151,948,366 | A/G  | 0.35 | 1.07 (1.05–1.09) 3.09 × 10<sup>-9</sup> | 0.06 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 |
| 2      | rs12173570        | Overall breast cancer (BCAC) | rs12173570 | 151,957,714 | T/C  | 0.10 | 1.12 (1.08–1.15) 6.46 x 10<sup>-10</sup> | 0.01 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 |
| 3      | rs851984          | Meta-analysis (BCAC ER<sup>-</sup> CIMBA) | rs851984 | 152,020,390 | C/A  | 0.41 | 1.08 (1.05–1.10) 9.65 x 10<sup>-12</sup> | 0.01 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 |
| 4      | rs9918437         | Overall breast cancer (BCAC) | rs9918437 | 152,072,718 | G/T  | 0.07 | 1.05 (1.01–1.09) 9.73 x 10<sup>-12</sup> | 0.01 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 |
| 5      | rs2747652         | Overall breast cancer (BCAC) | rs2747652 | 152,437,016 | T/C  | 0.54 | 1.05 (1.01–1.09) 9.73 x 10<sup>-12</sup> | 0.01 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 | 1.52 x 10<sup>8</sup> 3.67 x 10<sup>-15</sup> 0.89 |

Bold text highlights the top candidate SNPs within each signal. Freq., frequency.