

Large-scale genotyping identifies 41 new loci associated with breast cancer risk

- Kyriaki Michailidou,
- Per Hall,
- Anna Gonzalez-Neira,
- Maya Ghoussaini,
- Joe Dennis,
- Roger L Milne,
- Marjanka K Schmidt,
- Jenny Chang-Claude,
- Stig E Bojesen,
- Manjeet K Bolla,
- Qin Wang,
- Ed Dicks,
- Andrew Lee,
- Clare Turnbull,
- Nazneen Rahman,
- The Breast and Ovarian Cancer Susceptibility Collaboration,
- Olivia Fletcher,
- Julian Peto,
- Lorna Gibson,
- Isabel dos Santos Silva,
- Heli Nevanlinna,
- Taru A Muranen,
- Soo Hwang Teo,
- Cheng Har Yip,
- Sze Yee Phuah,
- Mikael Hartman,

Breast cancer is the most common cancer among women. Common variants at 27 loci have been identified as associated with susceptibility to breast cancer, and these account for ~9% of the familial risk of the disease. We report here a meta-analysis of 9 genome-wide association studies, including 10,052 breast cancer cases and 12,575 controls of European ancestry, from which we selected 29,807 SNPs for further genotyping. These SNPs were genotyped in 45,290 cases and 41,880 controls of European ancestry from 41 studies in the Breast Cancer Association Consortium (BCAC). The SNPs were genotyped as part of a collaborative genotyping experiment involving four consortia (Collaborative Oncological Gene-environment Study, COGS) and used a custom Illumina iSelect genotyping array, iCOGS, comprising more than 200,000 SNPs. We identified SNPs at 41 new breast cancer susceptibility loci at genome-wide significance ($P < 5 \times 10^{-8}$). Further analyses suggest that more than 1,000 additional loci are involved in breast cancer susceptibility.

Introduction

Breast cancer is the most commonly occurring malignancy among women, with an estimated 1 million new cases and over 400,000 deaths annually worldwide¹. Familial aggregation and twin studies have shown the substantial contribution of inherited susceptibility to breast cancer^{2,3}. Many genetic loci are known to contribute to this familial risk, including genes with high-penetrance mutations (notably *BRCA1* and *BRCA2*), moderate-risk alleles in genes such as *ATM*, *CHEK2* and *PALB2*, and common lower penetrance alleles, of which 27 have been identified so far, principally through genome-wide association studies (GWAS). In total, these loci explain approximately 30% of the familial risk of breast cancer¹⁵. Global analysis of GWAS data suggests that a substantial fraction of the residual aggregation can be explained by other common variants not yet identified, but the relative contributions of common and rare variants are still uncertain.

Results

To identify additional susceptibility loci for breast cancer, we first conducted a meta-analysis of 9 breast cancer GWAS in populations of European ancestry, including 10,052 cases and 12,575 controls ([Supplementary Table 1](#)). From this analysis, we selected 35,084 SNPs on the basis of evidence of association with breast cancer, derived from a 1-degree-of-freedom trend test, a test weighted for family history, a 2-degrees-of-freedom test and subset analyses based on cases of breast cancer diagnosed before 40 years of age and before 50 years of age (Online Methods). In particular, we were able to select all SNPs or surrogate SNPs with 1-degree-of-freedom $P_{\text{trend}} < 0.008$. To evaluate these SNPs, we then designed a custom Illumina iSelect genotyping array (iCOGS) in collaboration with three other consortia studying, in addition to breast cancer risk, susceptibility to ovarian cancer, prostate cancer and breast and ovarian cancers in *BRCA1* and *BRCA2* mutation carriers (COGS)^{17, 18, 19, 20}. The array included, in addition to SNPs selected from GWAS, SNPs selected for fine mapping of known susceptibility loci, functional candidate SNPs and SNPs related to other traits (Online Methods and [Supplementary Note](#)). The iCOGS array comprised 211,155 SNPs. These arrays were used to genotype 114,255 DNA samples from 52 studies participating in BCAC ([Supplementary Table 2](#)). After quality control exclusions (Online Methods and [Supplementary Table 3](#)), data were obtained for 199,961 SNPs in 52,675 cases and 49,436 controls. The analyses presented here are based on data from subjects of European ancestry (45,290 cases and 41,880 controls from 41 studies) and focus on 29,807 SNPs that were selected on the basis of the GWAS analysis that were successfully genotyped and were not located in regions previously known to be associated with breast cancer.

The association between each SNP and breast cancer risk was tested using a 1-degree-of-freedom trend test adjusted for study and seven principal components (Online Methods). There was some evidence for inflation in the test statistics, detected using data from 22,897 uncorrelated SNPs on iCOGS not selected on the basis of breast cancer risk ($\lambda = 1.20$, $\lambda_{1000} = 1.005$; [Supplementary Fig. 1a](#)). There was, however, clear evidence of an excess of statistically significant associations among the SNPs selected from the GWAS analysis ([Table 1](#) and [Supplementary Fig. 1b](#)). Although some excess was also observed among the SNPs not selected from the breast cancer GWAS, the excess of statistically significant associations was much more marked among the GWAS SNPs at all levels of statistical significance. In addition, of 21,128 SNPs not selected for breast cancer association that were also present in the combined GWAS data set, 10,864 (51%) had effects in the same direction in the GWAS and iCOGS data,

and, for these SNPs, inflation was 1.26 ($\lambda_{1000} = 1.007$) compared with 1.14 ($\lambda_{1000} = 1.0035$) for SNPs with effects in opposite directions in the two stages. A similar direction of effect was seen for these SNPs in the combined GWAS ($\lambda = 0.87$ for SNPs with effects in the same direction versus $\lambda = 0.79$ for SNPs with effects in the opposite direction, with inflation being <1 because SNPs showing evidence of association were excluded). Taken together, these results suggest that much of the inflation in the test statistics for SNPs not selected for breast cancer association is also due to the effect of true associations. Moreover, some of the excess of statistically significant associations seen in the SNPs not selected for breast cancer association was due to SNPs close to breast cancer-associated SNPs. For example, of the 45 SNPs with significant association at $P < 0.00001$, 21 were within 1 Mb of 1 of the newly identified breast cancer loci identified at our set genome-wide significance threshold. Taken together, these results strongly suggest that most of the excess of significant association for the GWAS-selected SNPs reflect true associations.

Table 1: Summary of SNPs by level of statistical significance in the iCOGS stage

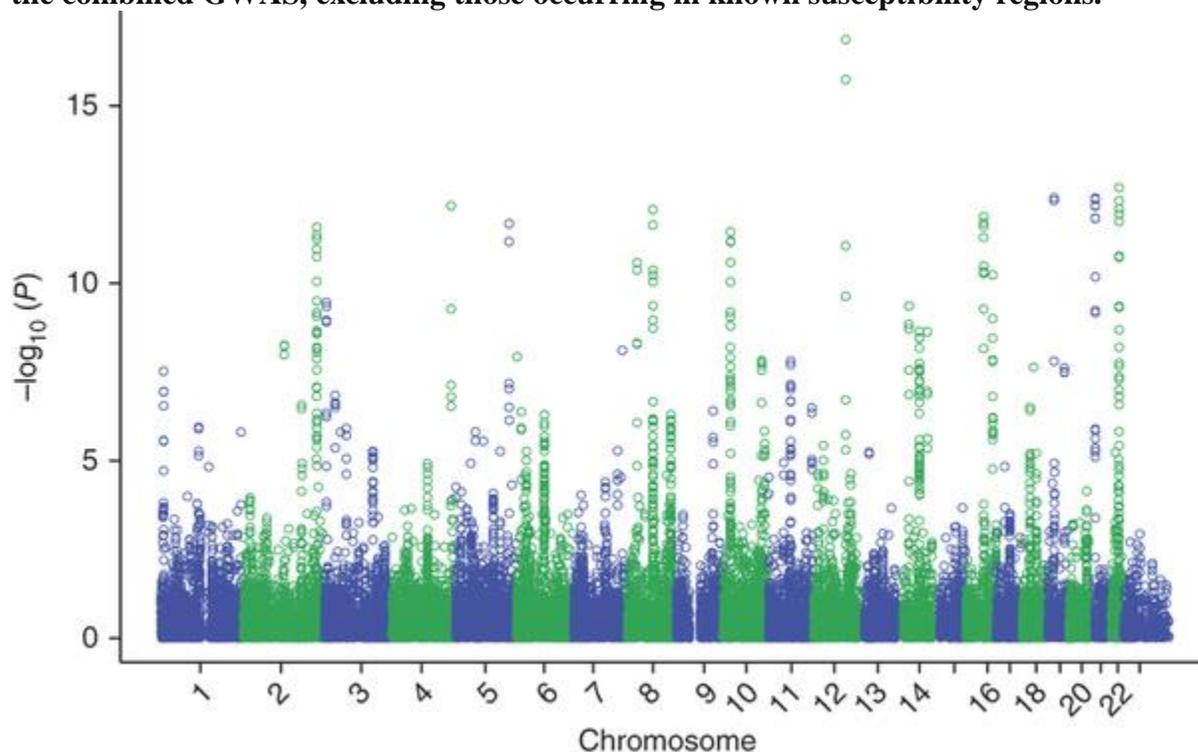
Of the 27 previously established breast cancer-associated loci, all but 4 showed clear evidence of association with overall breast cancer risk in the iCOGS stage ($P = 2.2 \times 10^{-5}$ – $P = 5.9 \times 10^{-125}$; [Supplementary Table 4](#)). Three loci showed weaker evidence for association: rs1045485, encoding an Asp302His variant in *CASP8*, whose association was previously identified in a candidate gene study ($P = 0.054$ in the iCOGS stage; $P = 0.0013$ in combined data from the GWAS and iCOGS stages)²¹; rs2380205 at 10p15, identified in a GWAS but suggested to be a possible false positive association in a previous BCAC analysis^{22,23} (iCOGS $P = 0.075$; combined $P = 0.0021$); and rs8170 at 19p13.1, for which the association has been shown to be specific to estrogen receptor (ER)-negative breast cancer²⁴ ($P = 0.0027$ in iCOGS; combined $P = 0.0012$). One locus, rs2284378 at 20q11, recently shown to be associated with ER-negative breast cancer, was not selected for the iCOGS array¹⁶.

Identification of new susceptibility loci

When the results from the GWAS and the iCOGS array were combined, 263 SNPs in 37 new regions had associations that reached $P < 5 \times 10^{-8}$ ([Fig. 1](#), [Table 2](#) and [Supplementary Figs. 2 and 3](#)). In four regions (5q11.2, 8q21.11, 10p12.31 and 18q11.2), this set of SNPs included SNPs within 1 Mb of each other that were uncorrelated, such that a second SNP was associated with disease after adjustment for the most significantly associated SNP ([Supplementary Fig. 4](#) and [Supplementary Table 5](#)). There was little or no evidence for heterogeneity in the per-allele odds ratios (ORs) among studies for any SNP (per-SNP² and P values are given in [Supplementary Fig. 2](#) and [Supplementary Table 6](#)). Genotype-specific OR estimates were consistent with a log-additive (allele dose) model for most SNPs, with the exception of three SNPs (rs616488, rs204247 and rs720475) for which the heterozygotes had a similar OR as homozygotes for the high-risk allele and two SNPs (rs11242675 and rs6472903) that were more consistent with a recessive model ([Supplementary Table 6](#)). Consistent with the pattern seen for previously established loci, there was strong evidence for specificity of the association to tumor subtype. For 13 of the loci, the per-allele OR was higher for ER-positive disease than for ER-negative disease (case-only $P < 0.05$), in most instances with little or no evidence of an association with ER-negative disease (based on data from 7,465 ER-negative cases and 27,074 ER-positive cases; [Supplementary Table 7a](#)). The most notable differences were for SNP

rs6828523 at 4q34.1 (ER-positive OR = 0.87 (95% confidence interval (CI) = 0.84–0.90); ER-negative OR = 1.01 (95% CI = 0.96–1.07); P for difference = 1.2×10^{-7}) and for rs7072776 at 10p12.31, where the estimated effects were in opposite directions (ER-positive OR = 1.09 (95% CI = 1.06–1.12); ER-negative OR = 0.94 (95% CI = 0.90–0.98); P for difference = 3.1×10^{-10}). No such difference was observed for the neighboring SNP rs11814448, which was associated with both ER-positive and ER-negative disease in the same direction. For one locus, SNP rs17817449 on chromosome 16, the association was stronger for ER-negative than for ER-positive disease (P for difference = 0.039). All SNPs showed comparable ORs for invasive and *in situ* disease (based on data from 2,335 ductal carcinoma *in situ*, DCIS, and 42,118 invasive cases), with the exceptions of rs12493607 and rs3903072, for which associations seemed to be restricted to invasive disease ([Supplementary Table 7b](#)). Two loci (rs2588809 at 14q24.1 ($P = 0.001$) and rs941764 at 14q32.12 ($P = 0.007$)) showed higher per-allele ORs for cases diagnosed at a young age ([Supplementary Table 7c](#)). Consistent with the predictions of a polygenic model of susceptibility²⁵, for 26 of the loci, the estimated OR was higher when restricted to cases with a positive family history for disease (significant at $P < 0.05$ for 5 loci), whereas for only 6 loci was the OR lower when restricted to cases with a positive family history ([Supplementary Table 7d](#)).

Figure 1: One-degree-of-freedom trend-test statistics for 29,807 iCOGS SNPs selected from the combined GWAS, excluding those occurring in known susceptibility regions.



The red horizontal line represents $P = 5 \times 10^{-8}$. The blue horizontal line represents $P = 1 \times 10^{-5}$.

Table 2: Results for 41 SNPs for which association $P < 5 \times 10^{-8}$ in combined GWAS and iCOGS analysis

Four of the newly associated loci (rs16857609 at 2q35, rs10759243 at 9q31, rs11199914 at 10q26 and rs2588809 at 14q24) lie close to regions previously associated with breast cancer risk. In each locus, however, the lead SNP was not correlated with the most strongly associated known association, and the association of the new SNP remained similarly statistically significant after adjustment for the previously associated SNP ([Supplementary Table 5](#)). In the case of rs2588809, which lies in *RAD51B* (also known as *RAD51LI*), the association was markedly stronger for ER-positive disease ($P = 0.011$; [Supplementary Table 7a](#)), whereas the previously associated SNPs (rs999737 and rs10483813), which lie ~370 kb telomeric, are associated with similar ORs for both ER-positive and ER-negative disease²⁶.

Two associated loci lie within or close to known breast cancer susceptibility genes. rs11571833 is a polymorphic variant in *BRCA2* that introduces a premature stop codon (p.Lys3326*), previously reported to have no association with breast cancer risk²⁷. The results from the current study, however, indicate that this variant is associated with a modestly higher risk of breast cancer. Further work will be required to determine whether this association is due to a higher risk variant or variants in linkage disequilibrium (LD). SNP rs132390 at 22q12 lies within an intron of *EMID1* but is ~500 kb upstream of *CHEK2*, raising the possibility that this association is mediated through the latter. *CHEK2* c.1100delC, the major deleterious *CHEK2* variant in European populations²⁸, occurs more frequently in association with the risk allele at rs132390 ($r^2 = 0.06$); however, the association between rs132390 and breast cancer risk persisted after adjustment for *CHEK2* c.1100delC, although attenuated (unadjusted OR in iCOGS = 1.12, $P = 5.9 \times 10^{-6}$; adjusted OR = 1.09, $P = 0.04$).

In addition to rs11571833, one further SNP is a coding variant: rs11552449 encodes a missense substitution p.His61Tyr in *DCLRE1B* (also known as *SNM1B*), an evolutionarily conserved gene involved in DNA stability and the repair of interstrand cross-links²⁹. The remaining loci are either intronic (20) or intergenic (19). Two loci lie within genes previously proposed as candidate breast cancer susceptibility genes. SNP rs12493607 lies in intron 2 of *TGFBR2*. An analysis of genes in the transforming growth factor (TGF)- β signaling pathway in European populations found weak evidence of an association between rs4522809 and breast cancer risk ($P = 0.02$)³⁰. This SNP is weakly correlated with rs12493607 ($r^2 = 0.25$) and also showed some evidence of association in our study, although weaker than that seen for rs12493607 (iCOGS $P = 0.00096$; combined analysis of GWAS and iCOGS $P = 0.0029$). A similar analysis of candidate SNPs in Asian populations identified SNP rs1078985 as a potential breast cancer susceptibility variant³¹. This variant, however, was uncorrelated with rs12493607 in Europeans and showed no evidence of association in our study ($P = 0.33$ in the iCOGS stage). SNP rs7904519 lies in intron 4 of *TCF7L2*. A previous candidate gene study found weak evidence for an association between a correlated SNP, rs12255372, associated with type 2 diabetes ($r^2 = 0.37$ with rs7904519), and familial breast cancer ($P = 0.04$)³².

The identification of the genes and variants underlying these associations will require more detailed fine mapping and functional analysis. Nevertheless, it is possible to discern some patterns. We identified 53 genes within 50 kb of the lead SNPs in the newly associated regions, totaling 96 genes when including the previously known loci. Analysis using Ingenuity Systems Pathway Analysis (IPA) identified an excess of genes reported to be involved in tumorigenesis (34 genes; $P = 0.0005$), breast cancer (15 genes; $P = 2 \times 10^{-5}$) and tumor incidence in model

systems (10 genes; $P = 2 \times 10^{-7}$). The most consistently over-represented functions were cell death ($P = 0.0028$), differentiation ($P = 2 \times 10^{-5}$) and expression ($P = 2 \times 10^{-8}$).

Three loci are located in the vicinity of susceptibility regions for other cancer types. SNP rs11780156 lies ~400 kb downstream of *MYC*. Previous GWAS have identified multiple loci upstream of *MYC* that are associated with different cancer types, including a locus for breast cancer. Functional studies have indicated that these associations might be mediated through transcriptional regulation of *MYC*. The newly associated locus is ~300 kb centromeric to a previously reported susceptibility locus for ovarian cancer, rs10088218, but is uncorrelated with it ($r^2 = 0.02$, based on data from European subjects in BCAC), raising the possibility that these loci might also be regulating *MYC*³³. SNP rs9790517 at 4q24 lies ~20 kb away from SNP rs7679673, previously reported to be associated with prostate cancer³⁴, and is correlated with it ($r^2 = 0.53$). SNP rs9790517 lies in intron 11 of *TET2*, which encodes a methylcytosine dioxygenase involved in myelopoiesis. Mutations in *TET2* are frequent in hematological malignancies but have also been reported in 2 of 47 breast tumors in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. In addition, Pharoah *et al.*¹⁸ have found an association between rs1243180 and ovarian cancer. This SNP is ~120 kb telomeric to rs7072776 and is partially correlated with it ($r^2 = 0.51$); both SNPs and the neighboring breast cancer-associated locus rs11814448 lie within the region 400 kb upstream of *DNAJC1*.

To further investigate the likely genes underlying the susceptibility variants, we examined associations between the lead SNPs and the RNA expression of neighboring genes in 473 primary breast tumors and 61 normal breast tissue samples in The Cancer Genome Atlas (TCGA) database. We found strong evidence for an association between rs616402 (a surrogate for rs616488; $r^2 = 0.66$) and expression of *PEX14* in both tumor ($P = 4.7 \times 10^{-12}$) and normal tissue ($P = 0.00018$; [Supplementary Table 8](#)), between rs3760983 (a surrogate for rs3760982; $r^2 = 1$) and expression of both *ZNF404* ($P = 1.2 \times 10^{-6}$ in tumors) and *ZNF283* ($P = 0.0089$) and between rs3903072 and expression of *CTSW* ($P = 4.9 \times 10^{-5}$). SNP rs3760982 was also found to be associated with the expression of *ZNF45* ($P = 0.0077$), *ZNF283* ($P = 0.05$) and *ZNF222* ($P = 0.01$) in lymphoblastoid cell lines from HapMap samples using the Genevar database³⁵ ([Supplementary Table 8c](#)). After adjustment for the SNP in the region most strongly associated with expression, SNP rs616488 and *PEX14* ($P = 0.0071$) as well as rs1217396 (a proxy for rs11552449) and *PTPN22* ($P = 0.0055$) and *DCLRE1B* ($P = 0.0067$) reached nominal significance at $P < 0.01$ ([Supplementary Table 8a](#)). Although none of these passed Bonferroni correction for multiple testing, the three associations found exceeded the number expected by chance with 46 associations tested. This supports some transcriptional effect from the risk-associated SNPs. *PEX14* is involved in peroxisome organization and protein and transmembrane transport; mutations in *PEX14* have been associated with Zellweger syndrome³⁶. The functions of *ZNF45*, *ZNF222* and *ZNF283* are unknown but may involve transcriptional regulation.

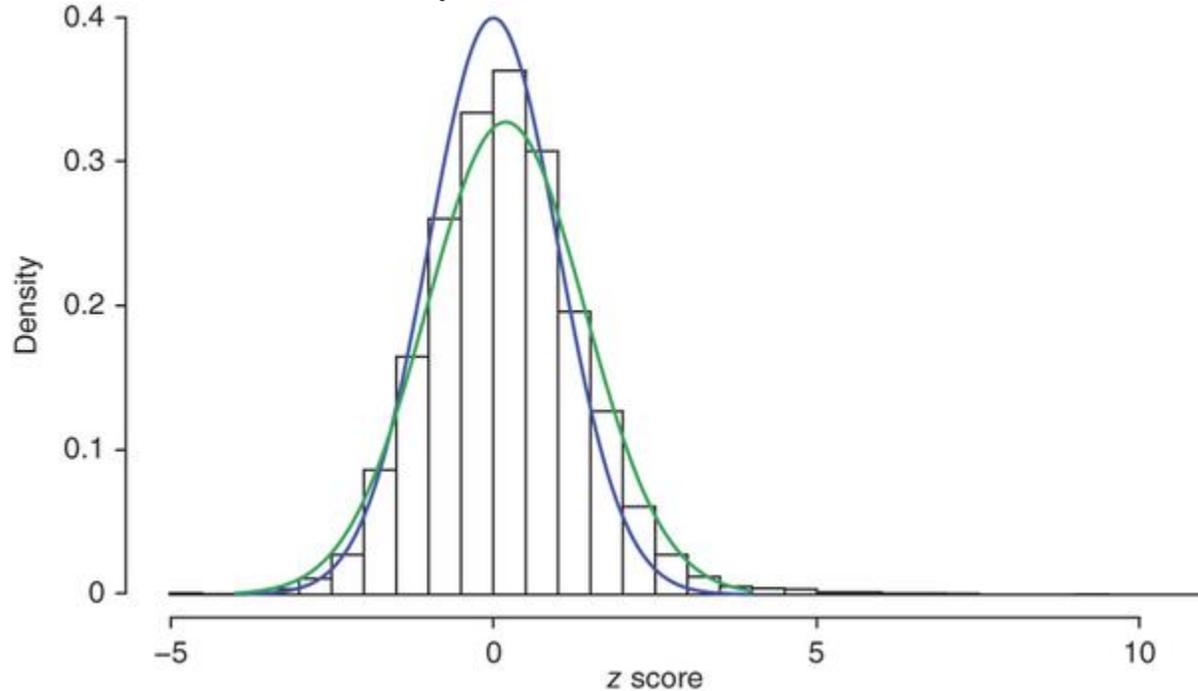
In addition to the genes described above, plausible candidate genes exist in several of the newly associated regions. *MUS81* at 11q13 has a key role in the maintenance of genomic stability and in DNA repair pathways^{37, 38}, and the cofilin gene (*CFL1*) is required for tumor cell motility and invasion, particularly in mammary tumors^{39, 40}. Several other genes have been associated with tumor aggressiveness; these include *PTH1R* at 3p21, *FOXQ1* at 6p25, *ARHGEF5* at 7q35 and *MKLI* at 22q13. *PTH1R* is the receptor for PTHLH, encoded by a previously identified

breast cancer susceptibility locus¹⁵. *PTHLH* is required for normal mammary gland function and has been shown to be involved in the metastasis of breast cancer cells to bone^{41, 42}. *FOXQ1* encodes a transcription factor with a key role in cell proliferation and migration and in breast cancer metastasis⁴³. Alterations in its expression level induce mesenchymal-epithelial transition⁴⁴. Dysfunctional *ARHGEF5* acts as an oncogene specific for human breast tissue, with a crucial role in tumorigenesis and metastasis in breast cancer⁴⁵. *MKLI* is also involved in tumor cell invasion and metastasis, particularly in human breast carcinoma⁴⁶. Two of the newly associated SNPs lie within the *TCF7L2* and *FTO* genes, previously associated with type 2 diabetes and/or obesity through GWAS^{47, 48, 49}. *TCF7L2* acts as a proto-oncogene and is involved in the Wnt pathway and in tumor formation⁵⁰. *PAX9* at 14q13.3 encodes a transcription factor that regulates cell proliferation, migration and resistance to apoptosis^{51, 52}. *SSBP4* is involved in DNA recombination and repair and has been suggested to have tumor suppressor activity^{53, 54}. The expression of *KREMEN1* at 22q12.1 is lower or absent in human tumors compared to normal tissue^{55, 56}. This gene encodes a negative regulator of the Wnt/ β -catenin pathway, which has a key role in cell fate determination, stem cell regulation and cell differentiation and proliferation. It has been suggested that lack of *KREMEN1* would activate the Wnt/ β -catenin pathway, thereby enhancing susceptibility to tumorigenesis^{55, 56}. Finally, *NTN4* at 12q22 encodes a secreted growth factor that regulates tumor growth. High levels of *NTN4* have been found in ER-positive but not ER-negative breast tumors⁵⁷. *NTN4* expression in tumors has also been suggested as a potential prognostic marker for breast cancer⁵⁷.

Overall contribution to breast cancer susceptibility

On the assumption that the risks conferred by common susceptibility loci combine multiplicatively (no interaction on a log-additive scale) and on the basis of the per-allele OR estimates from the iCOGS stage, we determined that the 41 newly associated loci explain approximately 5% of the familial risk of breast cancer. However, the overall excess of significant associations for SNPs selected from the breast cancer GWAS for genotyping in the iCOGS stage suggests that a much larger number of loci contribute to susceptibility, although they did not have associations reaching genome-wide levels of significance in the current study. To assess this hypothesis more formally, we identified a set of 10,668 SNPs selected from the GWAS that were uncorrelated ($r^2 < 0.1$ between any pair). Of these, the estimated OR was in the same direction as in the combined GWAS for 5,918 SNPs and in the opposite direction for 4,750 SNPs. Assuming that SNPs with effects in opposite directions are not associated with risk, an estimated 1,168 loci selected from the GWAS are associated with risk. However, this is an underestimate because weakly associated SNPs might have effects in opposite directions in the two stages. As an alternative approach, we fitted the distribution of z scores for the iCOGS stage, aligned to the direction of the effect in the GWAS, as a mixture of two normal distributions representing those SNPs that were or were not associated with disease ([Fig. 2](#) and Online Methods)⁵⁸. On the basis of the posterior probabilities from this analysis, an estimated 92% of loci ($n = 9,815$) were associated with breast cancer risk (95% CI = 85–100%), and these contributed approximately 18% of the familial risk of breast cancer. It should be noted, however, that the large majority of the loci had very small individual effects on risk: for example, the estimated OR was >1.05 for only 10 loci, and 920 loci had an estimated OR of >1.02 . When taking into account effects from the previously known loci, these analyses suggest that ~28% of familial risk is explained by common variants selected for iCOGS, of which ~14% can be explained by the 67 established loci (with a further ~20% due to higher penetrance loci).

Figure 2: Distribution of normalized effect sizes (z scores) in the iCOGS stage, with the direction of effect determined by the direction in the combined GWAS.



The blue curve represents the standard normal distribution. The green curve represents the best-fit normal distribution (mean = 0.19, s.d. = 1.22).

Discussion

To our knowledge, this is the largest genetic association study in cancer so far. The power of this approach is demonstrated by the fact that we have found evidence, at genome-wide levels of significance, for more than 40 new susceptibility loci, more than doubling the number of susceptibility loci for breast cancer. The effect sizes of the newly identified loci are generally modest (the highest OR was 1.26). However, the very high levels of statistical significance, the lack of heterogeneity among studies, the generally higher effect sizes for familial cases and the fact that most of the excess of significant associations was concentrated among SNPs selected on the basis of an association in the combined breast cancer GWAS all indicate that these are robust associations. Although the majority of the data are from populations of Northern and Western European ancestry, there was little or no evidence of heterogeneity in the OR estimates between studies, indicating that the associations apply broadly to populations of European ancestry. With more than 60 established breast cancer susceptibility loci, it is becoming possible to discern some more general patterns among the loci. Although most of the underlying genes and variants remain to be identified, there is a clear excess of genes either known to be involved in tumorigenesis in model systems or involved in processes relevant to cancer, such as cell death and differentiation. However, for other loci, such as *PEX14*, there is no obvious link to cancer susceptibility. Nine of the new loci lie in chromosomal regions with no known genes, suggesting that these may provide further examples of long-range regulation similar to that seen in the 8q24

region⁵⁹. We have identified three additional examples of loci in the vicinity of susceptibility loci for other cancers (*TET2*, 8q24 and *DNAJC1*). These associations might reflect the tissue-specific regulation of key genes, and understanding the functional mechanisms underlying these associations may be particularly informative.

On the basis of the current set of loci and assuming that all loci combine multiplicatively, the currently known loci now define a genetic profile for which 5% of the female population has a risk that is ~2.3-fold higher than the population average and for which 1% of the population has a risk that is ~3-fold higher. However, the large excess of significant associations among the SNPs selected from the GWAS suggests that many more susceptibility loci exist that have not met our threshold for genome-wide-significant association in this study and that these explain a similar fraction of the heritability as the currently known loci. The observation, made by comparing effect sizes in the iCOGS stage with those in the GWAS, that a very large number of loci, perhaps several thousand, contribute to polygenic susceptibility to breast cancer is consistent with results from GWAS in other complex disorders such as schizophrenia, using a different analytical approach⁶⁰. Incorporating these loci into risk models should substantially improve disease prediction, even if not all loci can be identified individually. Moreover, fine-scale mapping of the identified regions may uncover more of the missing heritability, either through identifying a more strongly associated variant (as found for the *CCND1* locus; see French *et al.*⁶¹) or by identifying additional signals (exemplified for the *TERT* region in Bojesen *et al.*⁶²). Genetic profiling using these common susceptibility loci in combination with rarer high-risk loci and other risk factors may provide a rational basis for targeted breast cancer preventi