

Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer

Genome-wide association studies (GWAS) and large-scale replication studies have identified common variants in 79 loci associated with breast cancer, explaining ~14% of the familial risk of the disease. To identify new susceptibility loci, we performed a meta-analysis of 11 GWAS, comprising 15,748 breast cancer cases and 18,084 controls together with 46,785 cases and 42,892 controls from 41 studies genotyped on a 211,155-marker custom array (iCOGS). Analyses were restricted to women of European ancestry. We generated genotypes for more than 11 million SNPs by imputation using the 1000 Genomes Project reference panel, and we identified 15 new loci associated with breast cancer at $P < 5 \times 10^{-8}$. Combining association analysis with ChIP-seq chromatin binding data in mammary cell lines and ChIA-PET chromatin interaction data from ENCODE, we identified likely target genes in two regions: *SETBP1* at 18q12.3 and *RNF115* and *PDZK1* at 1q21.1. One association appears to be driven by an amino acid substitution encoded in *EXO1*.

Breast cancer is the most common cancer in women worldwide¹. The disease aggregates in families and has an important inherited component. This inherited component is driven by a combination of rare variants, notably in *BRCA1*, *BRCA2*, *PALB2*, *ATM* and *CHEK2*, conferring a moderate or high lifetime risk of the disease, together with common variants at more than 70 loci identified through GWAS and large-scale replication studies^{2–20}. Taken together, these loci explain approximately one-third of the excess familial risk of breast cancer.

The majority of susceptibility SNPs have been identified through the Breast Cancer Association Consortium (BCAC), a collaboration involving more than 50 case-control studies. We recently reported the results of a large-scale genotyping experiment within BCAC, which used a custom array (iCOGS) designed to study variants of interest for breast, ovarian and prostate cancers. iCOGS comprised more than 200,000 variants, of which 29,807 had been selected from combined analysis of 9 breast cancer GWAS involving 10,052 breast cancer cases and 12,575 controls of European ancestry. In total, 45,290 breast cancer cases and 41,880 controls of European ancestry from 41 studies were genotyped with iCOGS, leading to the discovery of 41 new susceptibility loci¹⁶. A parallel analysis identified four loci specific to estrogen receptor (ER)-negative disease¹⁷. However, additional

susceptibility loci may have been missed because they were not selected from the original GWAS or were not included on the array.

Genotype imputation is a powerful approach to infer missing genotypes using the genetic correlations defined in a densely genotyped reference panel, thus providing the opportunity to identify new susceptibility variants even if they are not directly genotyped²¹. In this analysis, we aimed to identify additional breast cancer susceptibility loci by using data from all ~200,000 variants on the iCOGS array, and we used imputation to estimate genotypes for more than 11 million SNPs. We applied the same approach to data from 11 GWAS. After quality control exclusions, the data set comprised 15,748 breast cancer cases and 18,084 controls from GWAS together with 46,785 cases and 42,892 controls from 41 studies genotyped with iCOGS (Online Methods and Supplementary Table 1). All subjects were women of European ancestry.

We imputed genotypes using the 1000 Genomes Project March 2012 release as the reference data set (Online Methods). The main analyses were based on ~11.6 million SNPs that were imputed with imputation $r^2 > 0.3$ and had minor allele frequency (MAF) > 0.005 in at least one of the data sets²².

Of common SNPs (MAF > 0.05), 88% were imputed from the iCOGS array with $r^2 > 0.5$, compared with 99% of variants for the largest GWAS (UK2), which was genotyped using arrays with 594,375 SNPs (Fig. 1a,b and Supplementary Table 2)⁹. Of common SNPs, 37% were imputed on the iCOGS platform with $r^2 > 0.9$, in comparison to 85% for UK2. Thus, despite the iCOGS array being designed as a follow-up of GWAS for different diseases rather than a genome-wide array, the majority of common variants could be imputed using this array, but the overall imputation quality was poorer than from a standard GWAS array. Imputation quality decreased with decreasing allele frequency (Fig. 1c,d and Supplementary Table 2).

We calculated log odds ratio (OR) estimates and standard errors for each data set using logistic regression, adjusting for principal components when this adjustment was found to substantially reduce the inflation factor. We then combined the results from each data set for variants with MAF $> 0.5\%$ using a fixed-effects meta-analysis²³. We identified more than 7,000 variants with combined $P < 5 \times 10^{-8}$ for association, the large majority of which were in regions previously shown to be associated with breast cancer susceptibility. Of the 79 previously published breast cancer susceptibility loci identified in women of European ancestry, all but 8 showed evidence

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

Received 5 September 2014; accepted 11 February 2015; published online 9 March 2015; doi:10.1038/ng.3242

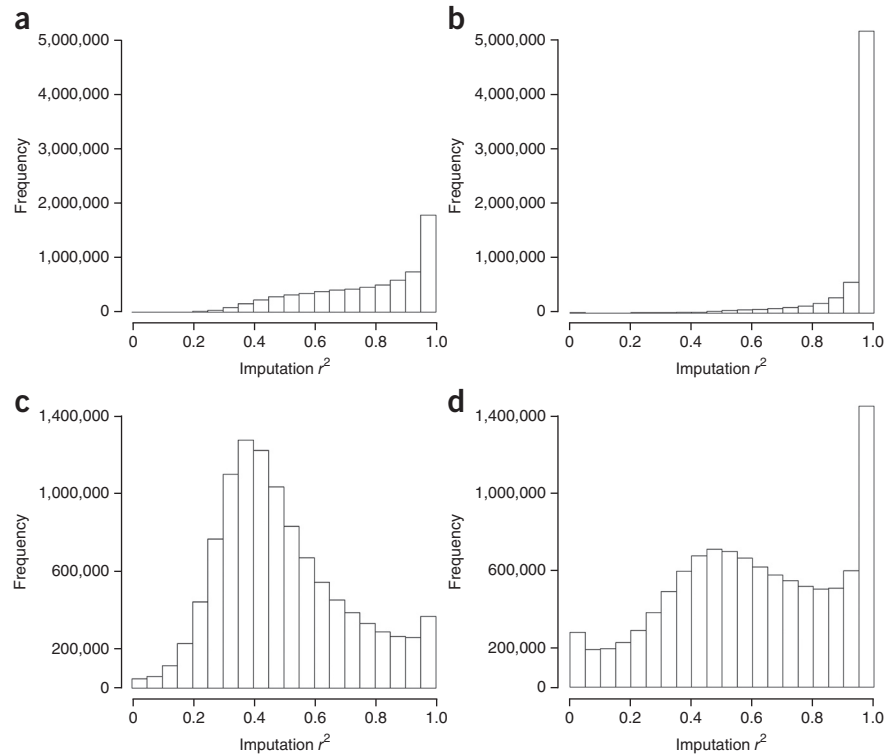
Figure 1 Histograms of imputation r^2 .

(a) Histogram of imputation r^2 for the iCOGS array for variants with MAF > 0.05.
 (b) Histogram of imputation r^2 for the UK2 GWAS for variants with MAF > 0.05.
 (c) Histogram of imputation r^2 for the iCOGS array for variants with MAF ≤ 0.05.
 (d) Histogram of imputation r^2 for the UK2 GWAS for variants with MAF ≤ 0.05.

of association at $P < 5 \times 10^{-8}$ for overall, ER-positive or ER-negative disease risk (**Supplementary Table 3**). For four of the eight variants (rs1550623 at 2q31, rs11571833 at 13q13.1, rs12422552 at 12p13.1 and rs11242674 at 6p25.3), we observed slightly weaker evidence of association. One reported variant, rs7726159, did not show association reaching $P < 5 \times 10^{-8}$ in this ($P = 0.0017$) or the previous analysis; it was identified through fine mapping of the *TERT* region at 5p15.33 (ref. 18). One other variant in *AKAP9*, rs6964587, reported previously¹⁹, did not show association reaching $P < 5 \times 10^{-8}$, but another variant correlated with it did ($P = 3.67 \times 10^{-8}$ for chr7:91681597:D; r^2 between the two markers = 0.98). The two remaining variants (rs2380205 at 10p15 and rs1045485 at *CASP8*) were reported in earlier analysis^{9,24} but did not have associations even reaching $P < 0.0001$, suggesting that these loci might have been false positive reports. An alternative variant at *CASP8*, rs1830298 ($r^2 = 0.06$, $D' = 1$ with rs1045485 in the 1000 Genomes Project CEU population (Utah residents of Northern and Western European ancestry)), did show association reaching $P < 5 \times 10^{-8}$ in this data set²⁵.

To assess evidence for additional susceptibility loci, we removed all SNPs within 500 kb of susceptibility variants identified previously in women of European ancestry^{2–14,16–19}, leaving 314 variants from 27 regions associated with breast cancer at $P < 5 \times 10^{-8}$ (**Supplementary Figs. 1 and 2**). We observed the strongest associations in a 610-kb interval (Build 37 coordinates 28,314,612–28,928,858) on chromosome 22 (smallest $P = 8.2 \times 10^{-22}$, for rs62237573). This interval lies approximately 100 kb centromeric to *CHEK2*, and further analysis showed that the associated SNPs were correlated with the *CHEK2* founder variant c.1100delC (strongest correlation $r^2 = 0.39$ for SNP rs62235635). *CHEK2* c.1100delC is known to be associated with breast cancer from candidate gene analysis but has not previously generated an association in GWAS^{26,27}. We performed an analysis adjusting for *CHEK2* c.1100delC using data on ~40,000 samples that had been genotyped for this variant. The strongest associated variant in this subset was rs140914118; after adjustment for c.1100delC, the statistical significance of the rs140914118 association was markedly diminished ($P = 3.1 \times 10^{-9}$ to $P = 0.78$; **Supplementary Fig. 3a,b**), suggesting that this signal is driven by *CHEK2* c.1100delC.

Variants in four regions (*DNAJC1*, 5p12, *PTHLH* and *MKL1*) lay within 2 Mb of a previously published breast cancer-associated SNP. In each case, these associations became weaker (no longer having $P < 5 \times 10^{-8}$) after adjustment for the previously associated SNP(s) in the region (data not shown). For four other regions, the significantly associated variants were identified in just one GWAS and failed imputation ($r^2 < 0.3$) in the remaining data sets, including iCOGS; we did not consider these variants further.



To confirm the results for the remaining 18 regions, we performed reimputation in the iCOGS data set without prephasing (Online Methods). Fifteen loci remained associated with breast cancer at $P < 5 \times 10^{-8}$ (**Table 1** and **Supplementary Table 4**). For three of the loci, the most significant SNP, or a highly correlated SNP, had been directly genotyped on iCOGS (**Supplementary Table 5**); one, rs11205277, had been included on the array because it is associated with adult height²⁸, whereas the other two were selected on the basis of evidence from the combined breast cancer GWAS but failed to reach genome-wide significance in the earlier analyses. We attempted to genotype the 12 remaining variants on a subset of ~4,000 samples to confirm the quality of the imputation (10 variants could be directly genotyped; for one region, an alternative, correlated variant was selected) (**Supplementary Table 5**). For the 11 variants that could be assessed, the r^2 estimates of correlation between the observed and imputed genotypes were close to the r^2 values estimated in imputation. Furthermore, the estimated effect sizes in the subset of individuals we genotyped were similar to those obtained from the imputed genotypes (**Supplementary Table 5**). These results indicate that the analyses based on imputed genotype data were reliable.

There was little or no evidence of heterogeneity in the per-allele OR estimates among the studies genotyped using iCOGS (**Supplementary Fig. 4** and **Supplementary Table 6**). There was little evidence for departure from a log-additive model for any locus, except for a borderline departure for rs6796502 ($P = 0.049$) for which the OR estimates for heterozygotes and homozygotes for the risk-associated allele were similar (**Supplementary Table 6**).

The estimated OR values for invasive versus *in situ* disease were similar for all the loci ($P > 0.05$) (**Supplementary Table 7**). For four of the loci (rs12405132, rs12048493, rs4593472 and rs6507583), the association was stronger for ER-positive disease (case-only $P < 0.05$) (**Supplementary Table 8**). Seven of the loci were associated with ER-negative disease ($P < 0.05$), but none had a stronger association for ER-negative than ER-positive disease. Two of the loci showed

Table 1 Results for the 15 newly discovered regions with combined $P < 5 \times 10^{-8}$

Best variant	Locus ^a	Position ^b	Alleles ^c	EAF ^d	r^2 ^e	GWAS OR (95% CI) ^f	GWAS P ^g	iCOGS OR (95% CI)	iCOGS P	Combined GWAS + iCOGS P	Genes within ± 2 kb	Enhancers in MCF-7 or HMEC cells	eQTLs
rs12405132	1q21.1	145,644,984	C/T	0.36	0.96	0.96 (0.92–0.99)	0.00962	0.95 (0.93–0.97)	2.34×10^{-7}	7.92×10^{-9}	<i>LOC10028814</i> , <i>RNF115</i> , <i>NBPF10</i> , <i>RNF115</i>	<i>RNF115</i> , <i>POLR3C</i> , <i>PDZK1</i> , <i>PIAS3</i>	–
rs12048493	1q21.2	149,927,034	A/C	0.34	0.76	1.04 (0.99–1.09)	0.121	1.07 (1.05–1.10)	1.66×10^{-9}	1.10×10^{-9}	–	–	–
rs72755295	1q43	242,034,263	A/G	0.03	0.94	1.19 (1.03–1.39)	0.021	1.15 (1.09–1.22)	2.60×10^{-7}	1.82×10^{-8}	<i>EXO1</i>	–	–
rs6796502	3p21.3	46,866,866	G/A	0.09	0.91	0.92 (0.87–0.98)	0.00657	0.92 (0.89–0.95)	8.13×10^{-7}	1.84×10^{-8}	–	–	–
rs13162653	5p15.1	16,187,528	G/T	0.45	0.72	0.92 (0.88–0.95)	5.18×10^{-6}	0.95 (0.93–0.97)	1.71×10^{-6}	1.08×10^{-10}	–	–	–
rs2012709	5p13.3	32,567,732	C/T	0.46	0.81	1.06 (1.02–1.09)	0.00101	1.05 (1.03–1.08)	1.66×10^{-6}	6.38×10^{-9}	–	–	–
rs7707921	5q14	81,538,046	A/T	0.23	0.88	0.94 (0.90–0.98)	0.00302	0.93 (0.91–0.95)	4.09×10^{-9}	5.00×10^{-11}	<i>ATG10</i>	–	<i>RPS23</i> , <i>ATP6AP1L</i>
rs9257408	6p22.1	28,926,220	G/C	0.38	0.92	1.05 (1.00–1.10)	0.0372	1.05 (1.03–1.08)	4.53×10^{-7}	4.84×10^{-8}	–	–	–
rs4593472	7q32.3	130,667,121	C/T	0.35	1.00	0.92 (0.88–0.96)	2.57×10^{-5}	0.95 (0.94–0.97)	3.97×10^{-6}	1.83×10^{-9}	<i>FLJ43663</i>	–	–
rs13365225	8p11.23	36,858,483	A/G	0.17	0.94	0.89 (0.85–0.93)	6.32×10^{-7}	0.95 (0.93–0.98)	0.000159	1.06×10^{-8}	–	–	–
rs13267382	8q23.3	117,209,548	G/A	0.36	0.97	1.07 (1.03–1.12)	0.000537	1.05 (1.03–1.07)	4.87×10^{-6}	1.72×10^{-8}	<i>LINC00536</i>	–	–
rs11627032	14q32.12	93,104,072	T/C	0.26	0.73	0.94 (0.90–0.98)	0.00114	0.94 (0.92–0.96)	1.06×10^{-6}	4.48×10^{-9}	<i>RIN3</i>	–	–
chr17:29230520:D	17q11.2	29,230,520	GGT/G	0.20	0.77	0.94 (0.89–0.98)	0.009	0.93 (0.91–0.96)	1.11×10^{-6}	3.34×10^{-8}	<i>ATAD5</i>	–	–
rs745570	17q25.3	77,781,725	A/G	0.50	0.93	0.94 (0.91–0.98)	0.000754	0.95 (0.93–0.97)	4.52×10^{-7}	1.40×10^{-9}	–	–	–
rs6507583	18q12.3	42,399,590	A/G	0.07	0.96	0.91 (0.85–0.98)	0.00803	0.91 (0.88–0.95)	1.21×10^{-6}	3.20×10^{-8}	<i>SETBP1</i>	–	–

Results are shown for the most strongly associated variant in the region.

^aChromosome. ^bBuild 37 position. ^cReference/effect alleles, based on the forward strand. ^dMean effect allele frequency over all controls. ^eImputation r^2 in the iCOGS samples (calculated by the average info score from IMPUTEv2). ^fPer-allele odds ratio for the minor allele relative to the major allele with the 95% confidence interval (CI). ^g P value for the 1-degree-of-freedom trend test.

significant trends in OR by age at diagnosis: for rs13162653, the OR was higher at younger ages ($P = 0.007$), whereas for rs6507583 the OR was higher at older ages ($P = 0.006$) (**Supplementary Table 9**). One of the variants (chr17:29230520:D in *ATAD5*) was correlated with a variant that has also been shown to be associated with serous ovarian cancer in a meta-analysis²⁹ ($r^2 = 0.93$ between chr17:29230520:D and chr17:29181220:I).

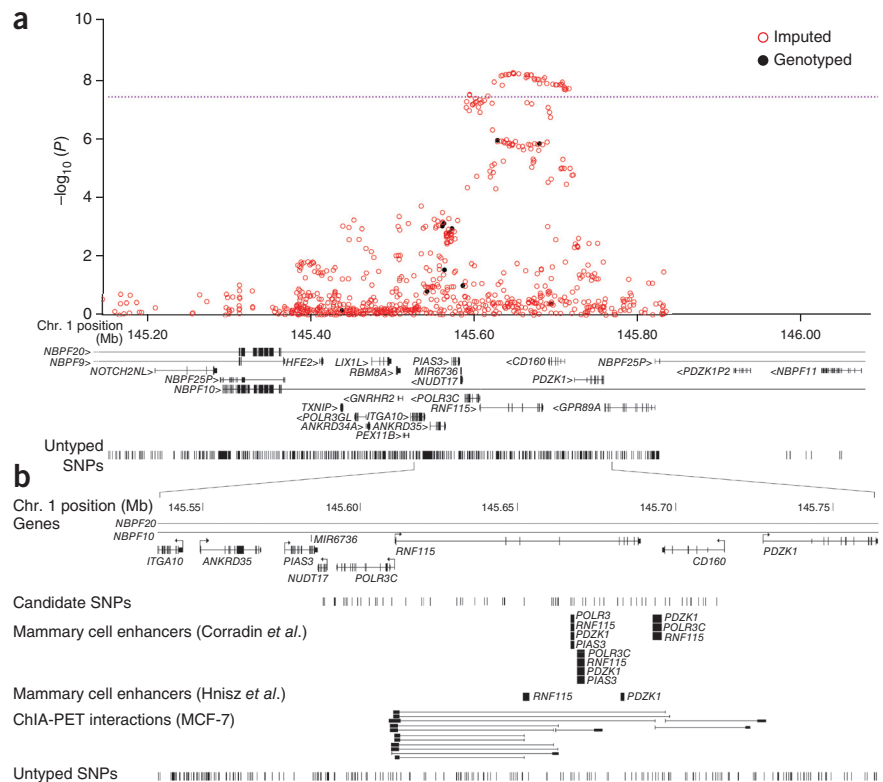
To approach the task of identifying the likely causal variants and genes underlying these associations, we first defined the set of all SNPs that were correlated with each of the 15 lead SNPs and that could not be ruled out as potentially causal (on the basis of a likelihood ratio of 100:1)³⁰, resulting in a subset of 522 variants (**Supplementary Table 10**). One of the variants, rs72755295, lies in an intron of *EXO1*, which encodes a protein involved in mismatch repair. This variant was strongly correlated with only one other variant, rs4149909, coding for an amino acid substitution in *EXO1* (p.Asn279Ser; with a Combined Annotation-Dependent Depletion (CADD) score of 33, indicating that it is likely to be deleterious)³¹, suggesting that this variant is likely to be functionally related to breast cancer risk. None of the remaining SNPs lay within gene-coding sequences, consistent with previous observations that most common cancer susceptibility variants are regulatory. For each of the remaining 520 variants, we then looked for enhancer elements in mammary cell lines, on the basis of Encyclopedia of DNA Elements (ENCODE) chromatin immunoprecipitation and sequencing (ChIP-seq) data^{32,33}. To identify potential gene targets, we combined this information with ENCODE chromatin interaction analysis with paired-end tag (ChIA-PET) chromatin interaction data. We identified two regions in which the associated variants overlapped with putative enhancer sequences and for which consistent promoter interactions were predicted (**Table 1**). For rs12405132 at 1q21.1, we identified four potential interacting genes: *RNF115*, *POLR3C*, *PDZK1* and *PIAS3* (**Fig. 2**). Of these, the strongest evidence was for *RNF115* and *PDZK1*; 3 of the 64 potentially causal variants lay in interacting enhancer regions. *RNF115* (also known as *BCA2*) encodes an E3 ubiquitin ligase RING finger protein that is overexpressed in ER-positive breast cancers³⁴. *PDZK1* encodes a scaffold protein that connects plasma membrane proteins and regulatory components, regulating their surface expression in epithelial cell apical domains, and has been proposed to act as an oncogene in breast cancer³⁵.

SNPs correlated with rs6507583 at 18q12.3 lay in regions interacting with the promoter of *SETBP1* (**Supplementary Fig. 5**). The encoded SETBP1 protein has been shown to bind the SET nuclear oncoprotein, which is involved in DNA replication.

We used data from The Cancer Genome Atlas (TCGA) to assess associations between the 15 newly discovered susceptibility variants and expression of neighboring genes in breast tumors and normal breast tissue. One SNP, rs7707921, was strongly associated with *RPS23* expression in all tissues (**Supplementary Fig. 6** and **Supplementary Table 11**). However, we observed stronger associations with expression for more telomeric SNPs that were less strongly associated with disease risk (top expression quantitative trait locus (eQTL) SNP rs3739: $P = 1 \times 10^{-23}$, P value for disease risk = 5.28×10^{-7}), suggesting that this association might be coincidental. SNP rs7707921 was also more weakly associated with the expression of *ATP6AP1L* ($P = 5.6 \times 10^{-5}$ in tumors, $P = 0.066$ in normal tissue).

On the basis of the estimated OR values in the iCOGS stage (all but one of which were in the range 1.05–1.10) and assuming that all loci combine multiplicatively, the 15 new loci identified here would explain a further ~2% of the twofold risk of breast cancer in the first-degree relatives of women with the disease. Taking the newly identified loci together with previously identified ones, more than 90

Figure 2 The chromosome 1 locus tagged by rs12405132. (a) The Manhattan plot shows the strength of genetic association versus chromosomal position, where each dot represents a genotyped or imputed SNP (in the iCOGS stage). The purple horizontal dotted line represents the threshold for genome-wide significance ($P = 5 \times 10^{-8}$). Gene structures are depicted as well as the locations of SNPs with MAF >0.01 that were neither imputed reliably nor genotyped. (b) Mammary cell enhancer locations as defined in Corradin *et al.*³² and Hnisz *et al.*³³ are shown, where elements overlapping the best associated SNPs are labeled with their predicted target genes. A subset of ChIA-PET interactions in MCF-7 cells (mediated by either RNA polymerase 2 or ER α) between enhancers and their target gene promoters are also shown.



independent common susceptibility loci for breast cancer have been identified, explaining ~16% of the familial risk. We estimate, assuming a log-additive model, that on the basis of genotypes for variants at these loci approximately 5% of women in the general population have >2-fold increased risk of breast cancer and 0.7% of women have >3-fold increased risk. In the current analyses, more than 50% of variants with MAF >0.005 in subjects of European ancestry could be imputed well ($r^2 > 0.5$). These results suggest that, although there may be further susceptibility variants with comparably associated effects that were not well imputed, the identification of many additional loci will require larger association studies. In the meantime, inclusion of these additional loci in polygenic risk scores will improve the ability to discriminate between individuals at high and low risk, potentially improving breast cancer screening and prevention.

URLs. Breast Cancer Association Consortium (BCAC), <http://apps.cce.medschl.cam.ac.uk/consortia/bcac/index.html>; Collaborative Oncological Gene-environment Study (COGS), <http://www.cogseu.org/>; ENCODE, <http://www.genome.gov/encode/> and <http://genome.ucsc.edu/ENCODE/>; iCOGS, <http://www.nature.com/icogs/> and <http://cce.medschl.cam.ac.uk/research/consortia/icogs/>; IMPUTE, <http://mathgen.stats.ox.ac.uk/impute/impute.html>; MACH, <http://csg.sph.umich.edu/abecasis/MACH/download/>; SHAPEIT, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; TCGA, <http://cancergenome.nih.gov/>; 1000 Genomes Project, <http://www.1000genomes.org/>.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

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

ACKNOWLEDGMENTS

The authors 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.

BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A12014) and by the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS). Meetings of the

BCAC have been funded by the European Union COST programme (BM0606). Genotyping on the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710, C8197/A16565), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer program and the Ministry of Economic Development, Innovation and Export Trade of Quebec, grant PSR-SIIRI-701. Combination of the GWAS data was supported in part by the US National Institutes of Health (NIH) Cancer Post-Cancer GWAS initiative, grant 1 U19 CA148065-01 (DRIVE, part of the GAME-ON initiative). For a full description of funding and acknowledgments, see the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

K. Michailidou and D.F.E. performed the statistical analysis and drafted the manuscript. D.F.E. conceived and coordinated the synthesis of the iCOGS array and led the BCAC. P.H. coordinated COGS. J. Benitez led the iCOGS genotyping working group. A.G.-N., G.P., M.R.A., N.A., D.H., J. Benitez, D.V., F.B., D.C.T., J.S., A.M.D., C.L., C. Baynes, S.A., C.S.H. and M.J.M. coordinated genotyping of the iCOGS array. M.G.-C., P.D.P.P. and M.K.S. led the BCAC pathology and survival working group. J.C.-C. led the BCAC risk factor working group. A.M.D. and G.C.-T. led the iCOGS quality control working group. J. Beesley, J.D. and M.J.L. provided bioinformatics support. M.K.B. and Q. Wang provided data management support for BCAC. S. Canisius provided analysis of the TCGA expression data. J.L.H., M.C.S., H.T. and C.A. coordinated ABCFS. M.K.S., A.B., S.V. and S. Cornelissen coordinated ABCS. K. Muir, A. Lophatananon, S.S.-B. and P.S. coordinated ACP. P.A.F., A. Hein, M.W.B. and L.H. coordinated BBCC. J.P., I.d.-S.-S., O.F. and L.G. coordinated BBCS. E.J.S., I.T., M.J.K. and N.M. coordinated BIGGS. P.K., D.J.H., S.L., S.M.G., M.M.G., W.R.D., C.A.H., F.S., B.E.H., L.L.M., C.D.B., S.J.C., J.F. and R.N.H. coordinated BPC3. B.B., F.M., H.S. and C. Sohn coordinated BSUCH. N.R. and C. Turnbull coordinated BOCs. P.G., T.T., C. Mulot and M. Sanchez coordinated CECILE. S.E.B., B.G.N., H.F. and S.F.N. coordinated CGPS. A.G.-N., J. Benitez, M.P.Z. and J.I.A.P. coordinated CNIO-BCS. H.A.-C. and S.L.N. coordinated CTS. H. Brenner, A.K.D., V.A. and C. Stegmaier coordinated ESTHER. A. Meindl, R.K.S., C. Sutter and R.Y. coordinated GC-HBOC. H. Brauch, U.H. and T.B. coordinated GENICA. H.N., T.A.M., K. Aittomäki, C. Blomqvist, K. Aaltonen and S.K. coordinated HEBCS. K. Matsuo, H. Ito, H. Iwata and K.T. coordinated HERPACC. T.D. and N.V.B. coordinated HMBCS. A. Lindblom and S. Margolin coordinated KARBAC. A. Mannermaa, V.K., V.-M.K. and J.M.H. coordinated KBCC. G.C.-T. and J. Beesley coordinated kConFab/AOCS. A.H.W., C. Tseng, D.V.D.B. and D.O.S. coordinated LAABC. D.L., P.N., H.W. and E.v.L. coordinated LMBC. J.C.-C., D.F.-J., U.E., S.B.

and A.R. coordinated MARIE. P.R., P. Peterlongo, S. Manoukian and L. Bernard coordinated MBCSG. F.J.C., J.E.O., E.H. and C.V. coordinated MCBCS. G.G.G., R.L.M. and C. McLean coordinated MCCS. C.A.H., B.E.H., F.S. and L.L.M. coordinated MEC. J.S., M.S.G., F.L. and M.D. coordinated MTLGEBCS. S.H.T., C.H.Y., N.A.M.T. and G.-H.T. coordinated MYBRCA. V.N.K., G.I.G.A. and S.N. coordinated NBCS. W.Z., S.L.H., M. Shrubsole and J. Long coordinated NBHS. R.W., K.P., A.J.-V. and M.G. coordinated OBCS. I.L.A., J.A.K., G.G. and A.M.M. coordinated OFBCR. P.D., R.A.E.M.T., C. Seynaeve and C.J.V.A. coordinated ORIGO. M.G.-C., J.F., S.J.C. and L. Brinton coordinated PBCS. K.C., H.D., M.E. and J.S.B. coordinated pKARMA. M.J.H., A. Hollestelle, J.W.M.M. and J.M.C. coordinated RBCS. P.H., J. Li, J. Liu and K.H. coordinated SASBAC. X.-O.S., W.L., Y.-T.G. and H.C. coordinated SBCGS. A.C., S.S.C. and M.W.R.R. coordinated SBCS. W.B., L.B.S. and Q.C. coordinated SCCS. M. Shah and B.J.P. coordinated SEARCH. D.K., J.-Y.C., S.K.P. and K.-Y.Y. coordinated SEBCS. M.H., H.M., K.S.C. and C.W.C. coordinated SGBCC. U.H., M.K. and D. Torres coordinated SKKDKFZS. A.J., J. Lubinski, K.J. and T.H. coordinated SZBCS. S. Sangrajrang, V.G., P.B. and J.M. coordinated TBCS. F.J.C., S. Slager, A.E.T., C.B.A. and D.Y. coordinated the TNBCC. C.-Y.S., C.-N.H., P.-E.W. and M.-F.H. coordinated TWBCS. A.J.S., A.A., N.O. and M.J.S. coordinated UKBGS. H.A., M.G.K., A.S.W., E.M.J., K.E.M., M.D.G., R.M.S., G.U., E.M., D.F.S. and G.C. coordinated EBCG. GWAS. Q. Waisfisz, H.M.-H., M.A.A. and R.B.v.d.L. coordinated DFBBCS. GWAS. D.F.E., N.R. and C. Turnbull coordinated UK2 GWAS. F.C., D. Trichopoulos, P. Peeters, E.L., M. Sund, K.-T.K., M.J.G., D.P., L.D., J.-M.H. and L.M.M. coordinated EPIC. All authors provided critical review of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

- Kamangar, F., Dores, G.M. & Anderson, W.F. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J. Clin. Oncol.* **24**, 2137–2150 (2006).
- Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
- Hunter, D.J. *et al.* A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**, 870–874 (2007).
- Stacey, S.N. *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **40**, 703–706 (2008).
- Stacey, S.N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **39**, 865–869 (2007).
- Ahmed, S. *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat. Genet.* **41**, 585–590 (2009).
- Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **41**, 324–328 (2009).
- Thomas, G. *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). *Nat. Genet.* **41**, 579–584 (2009).
- Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42**, 504–507 (2010).
- Antoniou, A.C. *et al.* A locus on 19p13 modifies risk of breast cancer in *BRCA1* mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat. Genet.* **42**, 885–892 (2010).
- Fletcher, O. *et al.* Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J. Natl. Cancer Inst.* **103**, 425–435 (2011).
- Haiman, C.A. *et al.* A common variant at the *TERT-CLPTM1L* locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.* **43**, 1210–1214 (2011).
- Ghoussaini, M. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat. Genet.* **44**, 312–318 (2012).
- Siddiq, A. *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum. Mol. Genet.* **21**, 5373–5384 (2012).
- Long, J. *et al.* Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet.* **8**, e1002532 (2012).
- Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361 (2013).
- Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.* **45**, 392–398 (2013).
- Bojesen, S.E. *et al.* Multiple independent variants at the *TERT* locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* **45**, 371–384 (2013).
- Milne, R.L. *et al.* Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Hum. Mol. Genet.* **23**, 6096–6111 (2014).
- Cai, Q. *et al.* Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat. Genet.* **46**, 886–890 (2014).
- Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
- Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Cox, A. *et al.* A common coding variant in *CASP8* is associated with breast cancer risk. *Nat. Genet.* **39**, 352–358 (2007).
- Lin, W.Y. *et al.* Identification and characterisation of novel associations in the *CASP8/ALS2CR12* region on chromosome 2 with breast cancer risk. *Hum. Mol. Genet.* **24**, 285–298 (2015).
- Meijers-Heijboer, H. *et al.* Low-penetrance susceptibility to breast cancer due to *CHEK2*(*)1100delC in noncarriers of *BRCA1* or *BRCA2* mutations. *Nat. Genet.* **31**, 55–59 (2002).
- CHEK2 Breast Cancer Case-Control Consortium. *CHEK2**1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am. J. Hum. Genet.* **74**, 1175–1182 (2004).
- Gudbjartsson, D.F. *et al.* Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–615 (2008).
- Kuchenbaecker, K.B. *et al.* Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat. Genet.* **47**, 164–171 (2015).
- Udler, M.S., Tyrer, J. & Easton, D.F. Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. *Genet. Epidemiol.* **34**, 463–468 (2010).
- Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
- Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).
- Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Wang, Z. *et al.* RNF115/BCA2 E3 ubiquitin ligase promotes breast cancer cell proliferation through targeting p21^{Waf1/Cip1} for ubiquitin-mediated degradation. *Neoplasia* **15**, 1028–1035 (2013).
- Kim, H. *et al.* PDZK1 is a novel factor in breast cancer that is indirectly regulated by estrogen through IGF-1R and promotes estrogen-mediated growth. *Mol. Med.* **19**, 253–262 (2013).

Kyriaki Michailidou¹, Jonathan Beesley², Sara Lindstrom³, Sander Canisius⁴, Joe Dennis¹, Michael J Lush¹, Mel J Maranian⁵, Manjeet K Bolla¹, Qin Wang¹, Mitul Shah⁵, Barbara J Perkins⁵, Kamila Czene⁶, Mikael Eriksson⁶, Hatf Darabi⁶, Judith S Brand⁶, Stig E Bojesen^{7–9}, Børge G Nordestgaard^{7–9}, Henrik Flyger¹⁰, Sune F Nielsen^{7,8}, Nazneen Rahman¹¹, Clare Turnbull¹¹, BOCS¹², Olivia Fletcher¹³, Julian Peto¹⁴, Lorna Gibson¹⁴, Isabel dos-Santos-Silva¹⁴, Jenny Chang-Claude¹⁵, Dieter Flesch-Janys^{16,17}, Anja Rudolph¹⁵, Ursula Eilber¹⁵, Sabine Behrens¹⁵, Heli Nevanlinna¹⁸, Taru A Muranen¹⁸, Kristiina Aittomäki¹⁹, Carl Blomqvist²⁰, Sofia Khan¹⁸, Kirsimari Aaltonen¹⁸, Habibul Ahsan^{21–25}, Muhammad G Kibriya^{21,22}, Alice S Whittemore^{26,27}, Esther M John^{26–28}, Kathleen E Malone²⁹, Marilie D Gammon³⁰, Regina M Santella³¹, Giske Ursin³², Enes Makalic³³, Daniel F Schmidt³³, Graham Casey³⁴, David J Hunter³, Susan M Gapstur³⁵, Mia M Gaudet³⁵, W Ryan Diver³⁵, Christopher A Haiman³⁴, Fredrick Schumacher³⁴, Brian E Henderson³⁴, Loic Le Marchand³⁶, Christine D Berg³⁷, Stephen J Chanock³⁸, Jonine Figueroa³⁸, Robert N Hoover³⁸, Diether Lambrechts^{39,40}, Patrick Neven⁴¹, Hans Wildiers⁴¹, Erik van Limbergen⁴¹, Marjanka K Schmidt⁴²,

Annegien Broeks⁴², Senno Verhoef⁴², Sten Cornelissen⁴², Fergus J Couch⁴³, Janet E Olson⁴⁴, Emily Hallberg⁴⁴, Celine Vachon⁴⁴, Quinten Waisfisz⁴⁵, Hanne Meijers-Heijboer⁴⁵, Muriel A Adank⁴⁵, Rob B van der Luijt⁴⁶, Jingmei Li⁶, Jianjun Liu⁴⁷, Keith Humphreys⁶, Daehae Kang^{48–50}, Ji-Yeob Choi^{49,50}, Sue K Park^{48–50}, Keun-Young Yoo⁵¹, Keitaro Matsuo⁵², Hidemi Ito⁵³, Hiroji Iwata⁵⁴, Kazuo Tajima⁵⁵, Pascal Guénel^{56,57}, Thérèse Truong^{56,57}, Claire Mulot⁵⁸, Marie Sanchez^{56,57}, Barbara Burwinkel^{59,60}, Frederik Marme^{59,61}, Harald Surowy^{59,60}, Christof Sohn⁵⁹, Anna H Wu³⁴, Chiu-chen Tseng³⁴, David Van Den Berg³⁴, Daniel O Stram³⁴, Anna González-Neira⁶², Javier Benitez^{62,63}, M Pilar Zamora⁶⁴, Jose Ignacio Arias Perez⁶⁵, Xiao-Ou Shu⁶⁶, Wei Lu⁶⁷, Yu-Tang Gao⁶⁸, Hui Cai⁶⁶, Angela Cox^{69,70}, Simon S Cross⁷¹, Malcolm W R Reed^{69,70}, Irene L Andrulis^{72,73}, Julia A Knight^{74,75}, Gord Glendon⁷², Anna Marie Mulligan^{76,77}, Elinor J Sawyer⁷⁸, Ian Tomlinson^{79,80}, Michael J Kerin⁸¹, Nicola Miller⁸¹, kConFab Investigators¹², AOCS Group¹², Annika Lindblom⁸², Sara Margolin⁸³, Soo Hwang Teo^{84,85}, Cheng Har Yip⁸⁵, Nur Aishah Mohd Taib⁸⁵, Gie-Hooi Tan⁸⁵, Maartje J Hooning⁸⁶, Antoinette Hollestelle⁸⁶, John W M Martens⁸⁶, J Margriet Collée⁸⁷, William Blot^{66,88}, Lisa B Signorello⁸⁹, Qiuyin Cai⁶⁶, John L Hopper⁹⁰, Melissa C Southey⁹¹, Helen Tsimiklis⁹¹, Carmel Apicella⁹⁰, Chen-Yang Shen^{92–94}, Chia-Ni Hsiung⁹³, Pei-Ei Wu^{92,93}, Ming-Feng Hou^{95,96}, Vessela N Kristensen^{97–99}, Silje Nord⁹⁷, Grethe I Grenaker Alnaes⁹⁷, NBCS¹², Graham G Giles^{90,100}, Roger L Milne^{90,100}, Catriona McLean¹⁰¹, Federico Canzian¹⁰², Dimitrios Trichopoulos^{89,103,104}, Petra Peeters^{105,106}, Eiliv Lund¹⁰⁷, Malin Sund¹⁰⁸, Kay-Tee Khaw¹⁰⁹, Marc J Gunter¹⁰⁶, Domenico Palli¹¹⁰, Lotte Maxild Mortensen¹¹¹, Laure Dossus^{112,113}, Jose-Maria Huerta¹¹⁴, Alfons Meindl¹¹⁵, Rita K Schmutzler^{116–118}, Christian Sutter¹¹⁹, Rongxi Yang^{59,60}, Kenneth Muir^{120,121}, Artitaya Lophatananon¹²⁰, Sarah Stewart-Brown¹²⁰, Pornthep Siriwanarangsarn¹²², Mikael Hartman^{123,124}, Hui Miao¹²⁴, Kee Seng Chia¹²⁴, Ching Wan Chan¹²⁵, Peter A Fasching^{126,127}, Alexander Hein¹²⁶, Matthias W Beckmann¹²⁶, Lothar Haeberle¹²⁶, Hermann Brenner^{128,129}, Aida Karina Dieffenbach^{128,129}, Volker Arndt¹²⁸, Christa Stegmaier¹³⁰, Alan Ashworth¹³, Nick Orr¹³, Minouk J Schoemaker¹¹, Anthony J Swerdlow^{11,131}, Louise Brinton³⁸, Montserrat Garcia-Closas^{11,13}, Wei Zheng⁶⁶, Sandra L Halverson⁶⁶, Martha Shrubsole⁶⁶, Jirong Long⁶⁶, Mark S Goldberg^{132,133}, France Labrèche^{134,135}, Martine Dumont^{136,137}, Robert Winqvist¹³⁸, Katri Pylkäs¹³⁸, Arja Jukkola-Vuorinen¹³⁹, Mervi Grip¹⁴⁰, Hiltrud Brauch^{129,141–143}, Ute Hamann¹⁴⁴, Thomas Brüning¹⁴⁵, GENICA Network¹², Paolo Radice¹⁴⁶, Paolo Peterlongo¹⁴⁷, Siranoush Manoukian¹⁴⁸, Loris Bernard^{149,150}, Natalia V Bogdanova¹⁵¹, Thilo Dörk¹⁵², Arto Mannermaa^{153–155}, Vesa Kataja^{156,157}, Veli-Matti Kosma^{153–155}, Jaana M Hartikainen^{153–155}, Peter Devilee^{158,159}, Robert A E M Tollenaar¹⁶⁰, Caroline Seynaeve¹⁶¹, Christi J Van Asperen¹⁶², Anna Jakubowska¹⁶³, Jan Lubinski¹⁶³, Katarzyna Jaworska¹⁶³, Tomasz Huzarski¹⁶³, Suleeporn Sangrajrang¹⁶⁴, Valerie Gaborieau¹⁶⁵, Paul Brennan¹⁶⁵, James McKay¹⁶⁵, Susan Slager⁴⁴, Amanda E Toland¹⁶⁶, Christine B Ambrosone¹⁶⁷, Drakoulis Yannoukakos¹⁶⁸, Maria Kabisch¹⁴⁴, Diana Torres^{144,169}, Susan L Neuhausen¹⁷⁰, Hoda Anton-Culver¹⁷¹, Craig Luccarini⁵, Caroline Baynes⁵, Shahana Ahmed⁵, Catherine S Healey⁵, Daniel C Tessier¹⁷², Daniel Vincent¹⁷², Francois Bacot¹⁷², Guillermo Pita⁶², M Rosario Alonso⁶², Nuria Álvarez⁶², Daniel Herrero⁶², Jacques Simard^{136,137}, Paul P D P Pharoah^{1,5}, Peter Kraft³, Alison M Dunning⁵, Georgia Chenevix-Trench², Per Hall⁶ & Douglas F Easton^{1,5}

¹Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ²Department of Genetics, QIMR (Queensland Institute for Medical Research) Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ³Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, Massachusetts, USA. ⁴Division of Molecular Carcinogenesis, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ⁵Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. ⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁷Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. ⁸Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. ⁹Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹⁰Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. ¹¹Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK. ¹²A full list of members and affiliations appears in the **Supplementary Note**. ¹³Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK. ¹⁴Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. ¹⁵Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁶Department of Cancer Epidemiology/Clinical Cancer Registry, University Clinic Hamburg-Eppendorf, Hamburg, Germany. ¹⁷Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany. ¹⁸Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. ¹⁹Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland. ²⁰Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. ²¹Center for Cancer Epidemiology and Prevention, University of Chicago, Chicago, Illinois, USA. ²²Department of Health Studies, University of Chicago, Chicago, Illinois, USA. ²³Department of Medicine, University of Chicago, Chicago, Illinois, USA. ²⁴Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. ²⁵Comprehensive Cancer Center, University of Chicago, Chicago, Illinois, USA. ²⁶Department of Health Research and Policy-Epidemiology, Stanford University School of Medicine, Stanford, California, USA. ²⁷Stanford Cancer

Institute, Stanford University School of Medicine, Stanford, California, USA. ²⁸Department of Epidemiology, Cancer Prevention Institute of California, Fremont, California, USA. ²⁹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ³⁰Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. ³¹Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York, USA. ³²Cancer Registry of Norway, Oslo, Norway. ³³Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Melbourne School of Population Health, Melbourne, Victoria, Australia. ³⁴Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. ³⁵Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA. ³⁶Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA. ³⁷Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins Medicine, Baltimore, Maryland, USA. ³⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA. ³⁹Vesalius Research Center, VIB, Leuven, Belgium. ⁴⁰Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium. ⁴¹Multidisciplinary Breast Center, University Hospitals Leuven, Leuven, Belgium. ⁴²Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ⁴³Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. ⁴⁴Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. ⁴⁵Department of Clinical Genetics, Section Oncogenetics, VU University Medical Center, Amsterdam, the Netherlands. ⁴⁶Division of Biomedical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands. ⁴⁷Human Genetics Division, Genome Institute of Singapore, Singapore. ⁴⁸Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. ⁴⁹Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea. ⁵⁰Cancer Research Institute, Seoul National University, Seoul, Korea. ⁵¹Seoul National University College of Medicine, Seoul, Korea. ⁵²Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka, Japan. ⁵³Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. ⁵⁴Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya, Japan. ⁵⁵Epidemiology Center for Disease Control and Prevention, Mie University Hospital, Tsu, Japan. ⁵⁶INSERM), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, Villejuif, France. ⁵⁷University Paris-Sud, UMRS 1018, Villejuif, France. ⁵⁸Université Paris Sorbonne Cité, UMRS 775, INSERM, Paris, France. ⁵⁹Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany. ⁶⁰Molecular Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁶¹National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany. ⁶²Human Genotyping (CEGEN) Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. ⁶³Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia, Spain. ⁶⁴Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid, Spain. ⁶⁵Servicio de Cirugía General y Especialidades, Hospital Monte Naranco, Oviedo, Spain. ⁶⁶Department of Medicine, Division of Epidemiology, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. ⁶⁷Shanghai Center for Disease Control and Prevention, Changning, China. ⁶⁸Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. ⁶⁹Sheffield Cancer Research Centre, University of Sheffield, Sheffield, UK. ⁷⁰Department of Oncology, University of Sheffield, Sheffield, UK. ⁷¹Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK. ⁷²Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. ⁷³Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ⁷⁴Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. ⁷⁵Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ⁷⁶Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. ⁷⁷Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada. ⁷⁸Research Oncology, Division of Cancer Studies, King's College London, Guy's Hospital, London, UK. ⁷⁹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. ⁸⁰Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. ⁸¹Department of Surgery, School of Medicine, National University of Ireland, Galway, Ireland. ⁸²Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. ⁸³Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. ⁸⁴Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Malaysia. ⁸⁵Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia. ⁸⁶Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands. ⁸⁷Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁸⁸International Epidemiology Institute, Rockville, Maryland, USA. ⁸⁹Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. ⁹⁰Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia. ⁹¹Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia. ⁹²Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ⁹³Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ⁹⁴School of Public Health, China Medical University, Taichung, Taiwan. ⁹⁵Cancer Center, Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung, Taiwan. ⁹⁶Department of Surgery, Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung, Taiwan. ⁹⁷Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Oslo, Norway. ⁹⁸Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ⁹⁹Department of Clinical Molecular Biology (EpiGen), University of Oslo, Oslo, Norway. ¹⁰⁰Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia. ¹⁰¹Department of Anatomical Pathology, Alfred Hospital, Melbourne, Victoria, Australia. ¹⁰²Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁰³Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece. ¹⁰⁴Hellenic Health Foundation, Athens, Greece. ¹⁰⁵Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, the Netherlands. ¹⁰⁶Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ¹⁰⁷Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Arctic University of Norway, Tromsø, Norway. ¹⁰⁸Department of Surgical and Perioperative Sciences, Umea University, Umea, Sweden. ¹⁰⁹School of Clinical Medicine, Cambridge Institute of Public Health, University of Cambridge, Cambridge, UK. ¹¹⁰Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute, Istituto per lo Studio e la Prevenzione Oncologica (ISPO), Florence, Italy. ¹¹¹Section for Epidemiology, Aarhus University, Aarhus, Denmark. ¹¹²INSERM Centre for Research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health Team, Villejuif, France. ¹¹³Université Paris-Sud, Villejuif, France. ¹¹⁴Department of Epidemiology, Consejería de Sanidad y Política Social, CIBER de Epidemiología y Salud Pública, Murcia, Spain. ¹¹⁵Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany. ¹¹⁶Center for Hereditary Breast and Ovarian Cancer, University Hospital Cologne, Cologne, Germany. ¹¹⁷Center for Integrated Oncology (CIO), University Hospital Cologne, Cologne, Germany. ¹¹⁸Center for Molecular Medicine Cologne (CMC), Faculty of Medicine, University of Cologne, Cologne, Germany. ¹¹⁹Institute of Human Genetics, University Hospital Heidelberg, Heidelberg, Germany. ¹²⁰Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, UK. ¹²¹Institute of Population Health, University of Manchester, Manchester, UK. ¹²²Ministry of Public Health, Nonthaburi, Thailand. ¹²³Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore. ¹²⁴Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore. ¹²⁵Division of General Surgery, National University Health System, Singapore. ¹²⁶Department of Gynecology and Obstetrics, University Breast Center Franconia, University Hospital Erlangen, Friedrich Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center of the Erlangen-Nuremberg Metropolitan Region, Erlangen, Germany. ¹²⁷Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. ¹²⁸Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹²⁹German Cancer Consortium (DKTK), Heidelberg, Germany. ¹³⁰Saarland Cancer Registry, Saarbrücken, Germany. ¹³¹Division of Breast Cancer Research, Institute of Cancer Research, London, UK. ¹³²Department of Medicine, McGill University, Montreal, Quebec, Canada. ¹³³Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, Quebec, Canada. ¹³⁴Département de Santé Environnementale et Santé au Travail, Ecole de Santé Publique, Université de Montréal, Montreal, Quebec, Canada. ¹³⁵Département de Médecine Sociale et Préventive, Ecole de Santé Publique, Université de Montréal, Montreal, Quebec, Canada. ¹³⁶Centre Hospitalier Universitaire de Québec Research Center, Quebec City, Quebec, Canada. ¹³⁷Department of Molecular Medicine, Laval University, Quebec City, Quebec, Canada. ¹³⁸Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, NordLab Oulu/Oulu University Hospital, Oulu, Finland. ¹³⁹Department of Oncology, Oulu University Hospital, University of Oulu, Oulu, Finland. ¹⁴⁰Department of Surgery, Oulu University Hospital, University of Oulu, Oulu, Finland. ¹⁴¹Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany. ¹⁴²University of Tübingen, Tübingen, Germany. ¹⁴³German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁴⁴Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁴⁵Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany. ¹⁴⁶Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico) Istituto Nazionale Tumori (INT), Milan, Italy. ¹⁴⁷Fondazione Istituto FIRC (Italian Foundation for Cancer Research) di Oncologia Molecolare (IFOM), Milan, Italy. ¹⁴⁸Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico) Istituto Nazionale Tumori (INT), Milan, Italy. ¹⁴⁹Department of Experimental Oncology, Istituto Europeo di Oncologia, Milan, Italy. ¹⁵⁰Cogentech Cancer Genetic Test Laboratory, Milan, Italy. ¹⁵¹Department of Radiation Oncology, Hannover Medical School, Hannover, Germany. ¹⁵²Gynaecology Research Unit, Hannover Medical School, Hannover, Germany.

¹⁵³School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland. ¹⁵⁴Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio, Finland. ¹⁵⁵Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland. ¹⁵⁶Cancer Center, Kuopio University Hospital, Kuopio, Finland. ¹⁵⁷Central Finland Hospital District, Jyväskylä Central Hospital, Jyväskylä, Finland. ¹⁵⁸Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. ¹⁵⁹Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands. ¹⁶⁰Department of Surgical Oncology, Leiden University Medical Center, Leiden, the Netherlands. ¹⁶¹Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, the Netherlands. ¹⁶²Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. ¹⁶³Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. ¹⁶⁴National Cancer Institute of Thailand, Bangkok, Thailand. ¹⁶⁵International Agency for Research on Cancer, Lyon, France. ¹⁶⁶Department of Molecular Virology, Immunology and Medical Genetics, Ohio State University, Columbus, Ohio, USA. ¹⁶⁷Roswell Park Cancer Institute, Buffalo, New York, USA. ¹⁶⁸Molecular Diagnostics Laboratory, Institute of Radioisotopes and Radiodiagnostic Products (IRRP), National Centre for Scientific Research 'Demokritos', Aghia Paraskevi Attikis, Athens, Greece. ¹⁶⁹Institute of Human Genetics, Pontificia Universidad Javeriana, Bogotá, Colombia. ¹⁷⁰Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, California, USA. ¹⁷¹Department of Epidemiology, University of California–Irvine, Irvine, California, USA. ¹⁷²McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. Correspondence should be addressed to D.F.E. (dfe20@medschl.cam.ac.uk).

ONLINE METHODS

Study overview. Details of the subjects, genotyping and quality control measures for the GWAS and iCOGS data are described elsewhere^{12,14,16,36,37}. All participating studies were approved by their appropriate ethics review boards, and all subjects provided informed consent. Analyses were restricted to women of European ancestry. All imputations were performed using the 1000 Genomes Project March 2012 release as the reference panel. Of the 11 GWAS, 8 (C-BCAC) plus a subset of the BPC3 GWAS (CGEMS) were used in the combined GWAS analysis that nominated 29,807 SNPs for the array. The BPC3 and TNBCC GWAS nominated additional SNPs with evidence for association with ER-negative or triple-negative (ER-, progesterone receptor (PR)- and HER2- negative) breast cancer. The EBCG GWAS was not used to nominate SNPs for the iCOGS array. For eight GWAS (C-BCAC), genotypes were imputed in a 2-stage procedure, using SHAPEIT to derive phased genotypes and IMPUTEv2 to perform imputation on the phased data²². We performed imputation using 5-Mb non-overlapping intervals for the whole genome. OR estimates and standard errors were obtained using logistic regression with SNPTTEST²¹. For two of the studies, we adjusted for the three leading principal components, as this adjustment was found to materially reduce the inflation factor; for the rest of the studies, no such adjustment was necessary. For the remaining three GWAS (BPC3, TNBCC and EBCG), imputation was performed using MACH and Minimac²³. Genomic control adjustment was applied to each GWAS as previously described¹⁶. The iCOGS data were also imputed in a two-stage procedure using SHAPEIT and IMPUTEv2, again using 5-Mb non-overlapping intervals. We split the ~90,000 samples into 10 subsets, keeping subjects from the same study in the same subset where possible. We obtained OR estimates and standard errors using logistic regression adjusting for study and nine principal components. For regions showing evidence of association, we repeated imputation in iCOGS, using IMPUTEv2 but without prephasing in SHAPEIT to improve imputation accuracy. We also increased the number of Markov chain Monte Carlo (MCMC) iterations from 30 to 90 and increased the buffer region from 250 kb to 500 kb.

Meta-analysis. OR estimates and standard errors were combined in a fixed-effects inverse variance meta-analysis using METAL²³. For the GWAS, results were included in the analysis for all SNPs with MAF >0.01 and imputation r^2 >0.3, except for the triple-negative GWAS, where the criteria were r^2 >0.9 and MAF >0.05. For iCOGS, we included all SNPs with r^2 ≥0.3 and MAF >0.005.

Confirmatory genotyping. The best variant in each region after reimputation and meta-analysis was genotyped in 4,123 samples from SEARCH, using TaqMan assays according to the manufacturer's instructions. Squared correlations between the observed genotypes and the genotypes estimated by imputation are shown in **Supplementary Table 5**. For all imputed SNPs,

the squared correlations were >0.7, the call rates were ≥0.98 and there was no evidence of departure of genotype frequencies from those expected under Hardy-Weinberg equilibrium (P > 0.1).

eQTL analyses. Germline genotype, mRNA expression and somatic copy number data for samples taken from breast tumors and tumor-adjacent normal tissue were obtained from the TCGA³⁸. Copy number and genotype data were obtained using the Affymetrix Genome-Wide Human SNP 6.0 platform. For mRNA expression data, we used the expression profiles obtained with the Agilent G4502A-07-3 microarray. Genotype data were subjected to the following quality control filters. SNPs were excluded in case of low frequency (MAF <1%), low call rate (<95%) or departure from Hardy-Weinberg equilibrium at P < 1×10^{-13} . Individuals were excluded on the basis of low call rate (<95%) or high heterozygosity (false discovery rate (FDR) < 1%). Furthermore, individuals were also excluded in case of non-European ancestry or male sex. Quality control and intersection with the other genomic data types resulted in 380 tumor samples and 56 normal samples.

Genotype data were imputed as described above. eQTL analysis was performed using linear regression with SNPTTEST, regressing the mRNA expression levels of selected candidate genes on the imputed genotype. For each gene, we performed eQTL analysis against every microarray probe that uniquely mapped to that gene. We adjusted the analyses for the somatic copy number of the gene and for SNPs that intersected the probe sequence, provided that the MAF for these SNPs exceeded 1% in individuals of European ancestry in the 1000 Genomes Project data.

Enhancer analyses. Maps of enhancer regions with predicted target genes were obtained from Hnisz *et al.*³³ and Corradin *et al.*³². Enhancers active in the mammary cell types MCF-7, HMEC and HCC1954 were intersected with candidate causal variants using Galaxy. ENCODE ChIA-PET chromatin interaction data from MCF-7 cells (mediated by RNA polymerase 2 and ERα) were downloaded using the UCSC Table browser. Galaxy was used to identify the ChIA-PET interactions between an implicated mammary cell enhancer (containing a strongly associated variant) and a predicted gene promoter (defined as regions 3 kb upstream and 1 kb downstream of the transcription start site).

36. Ahsan, H. *et al.* A genome-wide association study of early-onset breast cancer identifies *PFKM* as a novel breast cancer gene and supports a common genetic spectrum for breast cancer at any age. *Cancer Epidemiol. Biomarkers Prev.* **23**, 658–669 (2014).
37. Stevens, K.N. *et al.* 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. *Cancer Res.* **72**, 1795–1803 (2012).
38. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).