



Published in final edited form as:

Nat Genet. 2015 April ; 47(4): 373–380. doi:10.1038/ng.3242.

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

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

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to D.F.E. (df20@medschl.cam.ac.uk).

¹²A full list of members and affiliations appears in the Supplementary Note

Competing Financial Interests

The authors confirm that they have no competing financial interests

URLs

BCAC <http://ccge.medschl.cam.ac.uk/consortia/bcac/index.html>

COGS <http://www.cogseu.org/>

ENCODE <http://www.genome.gov/encode/>, genome.ucsc.edu/ENCODE/

iCOGS <http://www.nature.com/icogs/>, <http://ccge.medschl.cam.ac.uk/research/consortia/icogs/>

IMPUTE <http://mathgen.stats.ox.ac.uk/impute/impute.html>

MACH <http://www.sph.umich.edu/csg/abecasis/MACH/>

SHAPEIT https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html

TCGA <http://cancergenome.nih.gov/1000> genomes: <http://www.1000genomes.org/>

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 the Collaborative Oncological Gene-Environment Study (COGS). J. Benitez led the iCOGS genotyping working group. A.G.-N., G.P., M.R.A., J. Benitez, D.V., F.B., D.C.T., J. Simard, A.M.D., C.L., C. Baynes, S.A., C.S.H. and M.J.M. co-ordinated 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. co-ordinated ABCFS. M.K.S., A.B., S.V. and S. Cornelissen co-ordinated ABCS. K. Muir, A. Lophatananon, S.S.-B. and P.S. co-ordinated ACP. P.A.F., A.H., M.W.B. and L.H. co-ordinated BBCC. J.P., I.d.S.S., O.F. and L.G. co-ordinated BBBS. E.J.S., I.T., M.J.K. and N.M. co-ordinated 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.C., J.F. and R.N.H. co-ordinated BPC3. B.B., F.M., H.S. and C. Sohn co-ordinated BSUCH. P.G., T.T., C. Mulot and M. Sanchez co-ordinated CECILE. S.E.B., B.G.N., H.F. and S.F.N. co-ordinated CGPS. A.G.-N., J. Benitez, M.P.Z. and J.I.A.P. co-ordinated CNIO-BCS. H.A.-C. and S.L.N. coordinated CTS. H. Brenner, A.K.D., V.A. and C. Stegmaier co-ordinated ESTHER. A. Meindl, R.K.S., C. Sutter and R.Y. co-ordinated GC-HBOC. H. Brauch, U.H. and T.B. co-ordinated GENICA. H.N., T.A.M., K.A., C. Blomqvist, K.A. and S.K. co-ordinated HEBBS. K. Matsuo, H. Ito, H. Iwata and K.T. co-ordinated HERPACC. T.D. and N.V.B. co-ordinated HMBS. A. Lindblom and S. Margolin co-ordinated KARBAC. A. Mannermaa, V. Kataja, V.-M.K. and J.M.H. co-ordinated KBCP. G.C.-T. and J. Beesley co-ordinated kConFab/AOCS. A.H.W., C.-C.T., D.V.D.B. and D.O.S. co-ordinated 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. co-ordinated MARIE. P.R., P.P., S. Manoukian and L. Bernard co-ordinated MBCSG. F.J.C., J.E.O., E.H. and C.V. co-ordinated MCBCS. G.G.G., R.L.M. and C. McLean co-ordinated MCCS. C.A.H., B.E.H., F.S. and L.L.M. co-ordinated MEC. J. Simard, M.S.G., F.L. and M.D. co-ordinated MTLGEBBS. S.H.T., C.H.Y., Y.-C.T. and N.A.M.T. co-ordinated MYBRCA. V. Kristensen, G.I.G.A., S.N. and A.-L.B.-D. co-ordinated NBCS. W.Z., S.L.H., M. Shrubsole and J. Long coordinated NBHS. R.W., K.P., A.J.-V. and M.G. co-ordinated OBBS. 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. co-ordinated ORIGO. M.G.-C., J.F., S.J.C. and L. Brinton co-ordinated PBCS. K.C., H.D., M.E. and J. Brand co-ordinated pKARMA. J.W.M.M. and J.M.C. co-ordinated RBCS. P. Hall, J. Li, J. Liu and K.H. co-ordinated SASBAC. X.-O.S., W.L., Y.-T.G. and H.C. co-ordinated SBCGS. A.C., S.S.C. and M.W.R. Reed co-ordinated SBCS. W.B., L.B.S. and Q.C. coordinated SCCS. M. Shah and B.J.B. co-ordinated SEARCH. D.K., J.-Y.C., S.K.P. and K.-Y.Y. co-ordinated SEBBS. M.H., H.M., K.S.C. and C.W.C. co-ordinated SGBCC. U.H., M. Kabisch and D. Torres coordinated SKKDKFZS. A.J., J. Lubinski, K.J. and T.H., co-ordinated SZBBS. S. Sangrajrang, V.G., P.B. and J.M. co-ordinated TBBS. F.J.C., S. Slager, A.E.T., C.B.A. and D.Y. co-ordinated the TNBBS. C.-Y.S., C.-N.H., P.-E.W. and M.-F.H. co-ordinated TWBBS. A.S., A.A., N.O. and M.J.S. co-ordinated UKBBS. 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. co-ordinated EBCG GWAS. Q.W., H.M.-H., M.A.A. and R.B.v.d.L. co-ordinated DFBBBS GWAS. D.F.E., N.H. and C.T. co-ordinated UK2 GWAS. F.C., D. Trichopoulos, P.P., 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.

Abstract

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 of 15,748 breast cancer cases and 18,084 controls, and 46,785 cases and 42,892 controls from 41 studies genotyped on a 200K custom array (iCOGS). Analyses were restricted to women of European ancestry. Genotypes for more than 11M SNPs were generated by imputation using the 1000 Genomes Project reference panel. We identified 15 novel loci associated with breast cancer at $P < 5 \times 10^{-8}$. Combining association analysis with ChIP-Seq data in mammary cell lines and ChIA-PET chromatin interaction data in ENCODE, we identified likely target genes in two regions: *SETBP1* on 18q12.3 and *RNF115* and *PDZK1* on 1q21.1. One association appears to be driven by an amino-acid substitution 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 has 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 utilised 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 nine 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 novel susceptibility loci¹⁶. A parallel analysis identified four loci specific to oestrogen receptor (ER)-negative disease¹⁷. However, additional susceptibility loci may have been missed because they were not selected from the original GWAS, or 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 novel susceptibility variants even if not directly genotyped²¹. In this analysis we aimed to identify additional breast cancer susceptibility loci by utilising data from all 200k variants on the iCOGS array, and used imputation to estimate genotypes for more than 11M SNPs. We applied the same approach to data from 11 GWAS. After quality control (QC) exclusions, the dataset comprised 15,748 breast cancer cases and 18,084 controls from GWAS, and 46,785 cases and 42,892 controls from 41 studies genotyped with iCOGS (see Online Methods and Supplementary Tables 1a–1e). All subjects were women of European ancestry.

We imputed genotypes using the 1000 Genomes Project March 2012 release as the reference dataset (see Online Methods) The main analyses were based on ~11.6M SNPs that were imputed with imputation $r^2 > 0.3$ and had $MAF > 0.005$ in at least one of the datasets²².

Of common SNPs ($MAF > 0.05$), 88% were imputed from the iCOGS array with $r^2 > 0.5$; this compared to 99% of variants for the largest GWAS (UK2), which was genotyped using a 670k SNP array (Figure 1a and 1b, Supplementary Table 2). Thirty-seven per cent of common SNPs were imputed on the iCOGS with $r^2 > 0.9$, compared with 85% for UK2. Thus, despite 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 the iCOGS, but the overall imputation quality was, poorer than from a standard GWAS array. Imputation quality decreased with decreasing allele frequency (Figure 1c and 1d, Supplementary Table 2).

Log odds ratio estimates and standard errors were calculated for each dataset using logistic regression, adjusting for principal components where it was found to reduce substantially the inflation factor. We then combined the results from each dataset for variants with $MAF > 0.5\%$ using a fixed effects meta-analysis²³. More than 7,000 variants with a combined $P < 5 \times 10^{-8}$ for association were identified, the large majority of which was 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 eight show evidence of association at $P < 5 \times 10^{-8}$ for overall, ER-positive or ER-negative disease risk (Supplementary Tables 3a, 3b and 3c). For four of the eight variants, (rs1550623 on 2q31, rs11571833 on 13q13.1, rs12422552 on 12p13.1 and rs11242674 on 6p25.3), slightly weaker evidence of association was observed. One reported variant, rs7726159 did not reach $P < 5 \times 10^{-8}$ in this ($P = 0.0017$) or the previous analysis – it was identified through fine-mapping of the *TERT* region on 5p15.33¹⁸. One other variant in *AKAP9*, rs6964587 reported previously¹⁹ did not reach $P < 5 \times 10^{-8}$ but an alternative 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 on 10p15 and rs1045485 at *CASP8*) were reported in earlier analysis^{9,24} but did not even reach $P < 0.0001$, suggesting that they may have been false positive reports. An alternative variant at *CASP8*, rs1830298 ($r^2 = 0.06$, $D' = 1$ with rs1045485 in 1000G CEU) did reach $P < 5 \times 10^{-8}$ in this dataset²⁵.

To assess evidence for additional susceptibility loci, we removed all SNPs within 500kb 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 Figures 1 and 2). The strongest associations were observed in a 610kb (b37 28,314,612- 28,928,858) interval on chromosome 22 (smallest $P = 8.2 \times 10^{-22}$, for rs62237573). This interval lies approximately 100kb centromeric to *CHEK2*, and further analysis revealed that the associated SNPs were correlated with the *CHEK2* founder variant 1100delC (strongest correlation $r^2 = 0.39$ for SNP rs62235635), *CHEK2* 1100delC is known to be associated with breast cancer through candidate gene analysis, but has not previously generated an association in GWAS^{26,27}. We performed an analysis adjusting for *CHEK2* 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 1100delC

the statistical significance diminished markedly ($P=3.1\times 10^{-9}$ to $P=0.78$; Supplementary Figures 3a and 3b), suggesting that this signal is driven by *CHEK2* 1100delC.

Variants in four regions (*DNAJ1*, 5p12, *PTHLH* and *MKLI*) lay within 2Mb of a previously published susceptibility-associated SNP. In each case, these associations became weaker (no longer $P<5\times 10^{-8}$) after adjustment for the previously associated SNP(s) in the region (data not shown). For four other regions, the significant variants were identified in just one GWAS, and failed imputation ($r^2<0.3$) in the remaining datasets, including iCOGS; we did not consider these variants further.

To confirm the results for the remaining 18 regions, we performed re-imputation in the iCOGS dataset without phasing (See 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, while the other two were selected based on 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 ~4K 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 between the observed and imputed genotypes were close to the r^2 estimated in the imputation. Furthermore, the estimated effect sizes in the subset of individuals that 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 odds ratios (ORs) among studies genotyped using iCOGS (Supplementary Table 6 and Supplementary Figure 4). 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 ORs for heterozygotes and homozygotes for the risk associated allele were similar (Supplementary Table 6).

The estimated ORs 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 significant trends in the OR by age at diagnosis: for rs13162653, the OR was higher at younger ages ($P=0.007$), while for rs6507583, the OR was higher at older ages ($P=0.006$) (Supplementary Table 9). One of the variants, chr17:29230520:D in *ATAD5* is 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 correlated with each of the 15 lead SNPs and that could not be ruled out as potentially causal (based on a likelihood ratio 100:1³⁰),

resulting in a subset of 522 variants (Supplementary Table 10). One of the variants, rs72755295, lies in an intron of *EXO1*, encoding a protein involved in mismatch repair. It is strongly correlated with only one other variant, rs4149909, coding for an amino-acid substitution in *EXO1* (p.Asn279Ser; CADD score 33³¹), 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, based on ENCODE ChIP-Seq data^{32,33}. To identify potential gene targets, we combined this information with ENCODE 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* (Figure 2). Of these, the strongest evidence was for *RNF115* and *PDZK1*; three of the 64 potentially causal variants lay in interacting enhancer regions. *RNF115* (also known as *BCA2*) is an E3 ubiquitin ligase RING finger protein that is overexpressed in ER-positive breast cancers³⁴. *PDZK1* is a scaffold protein that connects plasma membrane proteins and regulatory components, regulating their surface expression in epithelial cells 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 Figure 5). The encoded protein has been shown to bind the SET nuclear oncogene which is involved in DNA replication.

We utilised data from TCGA to assess associations between the 15 novel susceptibility variants and expression of neighbouring genes in breast tumors and normal breast tissue. One SNP, rs7707921, was strongly associated with *RPS23* expression in all tissues (Supplementary Table 11, Supplementary Figure 6). However, stronger associations with expression were observed with more telomeric SNPs that were less strongly associated with disease risk (top eQTL SNP rs3739: $P=10^{-23}$, $P\text{-risk}=5.28\times 10^{-7}$), suggesting that this association may be coincidental. SNP, rs7707921 was also more weakly associated with expression of *ATP6AP1L* ($P=5.6\times 10^{-5}$ in tumours, $P=0.066$ in normal tissue).

Based on the estimated ORs in the iCOGS stage (all but one of which were in the range 1.05–1.10), the 15 novel loci identified here would explain a further ~2% of the 2-fold familial risk of breast cancer. Taken together with previously identified loci, more than 90 independent common susceptibility loci for breast cancer have been identified, explaining ~16% of the familial risk. We estimate assuming a log-additive model that, based on genotypes for variants at these loci, approximately 5% of women in the general population have a >2 fold increased risk of breast cancer and 0.7% of women have a >3 fold increased risk. In the current analyses, more than 50% of variants with MAF>0.005 in subjects of European ancestry were well imputable ($r^2>0.5$) These results suggest that, while there may be further susceptibility variants with comparable 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 our

ability to discriminate between high and low risk individuals, potentially improving breast cancer screening and prevention.

Online Methods

Details of the subjects, genotyping and QC 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 board 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-, 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 two-stage procedure, using SHAPEIT to derive phased genotypes and IMPUTEv2 to perform the imputation on the phased data²². We performed the imputation using 5Mb non-overlapping intervals for the whole genome. OR estimates and standard errors were obtained using logistic regression with SNPTEST²¹. For two of the studies we adjusted for the 3 leading principal components as it was found to reduce materially 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 5Mb non-overlapping intervals. We split the ~90K samples into 10 subsets, where possible keeping subjects from the same study in the same subset. We obtained OR estimates and standard errors using logistic regression adjusting for study and 9 principal components.

For the regions showing evidence of association we repeated the imputation in iCOGS, using IMPUTEv2 but without pre-phasing in SHAPEIT to improve imputation accuracy. We also increased the number of MCMC iterations from 30 to 90, and increased the buffer region from 250kb to 500kb.

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 TN GWAS where the criteria were $r^2>0.9$ and MAF>0.05. For iCOGS, we included all SNPs with $r^2\geq 0.3$ and MAF>0.005.

Confirmatory genotyping

The best variant in each region after the re-imputation and meta-analysis was genotyped in 4123 samples from SEARCH, using Taqman according to the manufacturer's instructions. The squared correlations between the observed genotypes and the genotypes estimated by imputation are shown in Supplementary Table 5. For all the imputed SNPs the squared

correlations was greater than 0.7, the call-rates were ≥ 0.98 and there was no evidence of departure of genotype frequencies from those expected under HWE ($p > 0.1$).

eQTL analyses

Germline genotype, mRNA expression, and somatic copy number data for samples taken from breast tumours and tumour-adjacent normal tissue were obtained from The Cancer Genome Atlas³⁸. The copy number and genotype data were measured using the Affymetrix Genome-Wide Human SNP 6.0 platform. For the mRNA expression data, we used the expression profiles obtained using the Agilent G4502A-07-3 microarray. The 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 $P < 1 \times 10^{13}$. Individuals were excluded based on low call rate ($< 95\%$), or high heterozygosity (false discovery rate $< 1\%$). Furthermore, individuals were also excluded in case of non-European ancestry, or male gender. Quality control and intersection with the other genomic data types resulted in 380 tumour samples and 56 normal samples.

The genotype data were imputed as described above. eQTL analysis was performed using linear regression with SNPTEST, regressing the mRNA expression of selected candidate genes on the imputed genotype. For each gene, we performed the eQTL analysis against every microarray probe that uniquely maps to that gene. We adjusted the analyses for somatic copy number of the gene, and for SNPs that intersect the probe sequence, provided that their MAF exceeds 1% in individuals of European ancestry in the 1,000 Genomes 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 MCF7, HMEC and HCC1954 were intersected with candidate causal variants using Galaxy. ENCODE ChIA-PET chromatin interaction data from MCF7 cells (mediated by RNApolII and ER α) were downloaded using the UCSC Table browser. Galaxy was used to identify 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).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Kyriaki Michailidou¹, Jonathan Beesley², Sara Lindstrom³, Sander Canisius⁴, Joe Dennis¹, Michael Lush¹, Mel J Maranian⁵, Manjeet K Bolla¹, Qin Wang¹, Mitul Shah⁵, Barbara J Perkins⁵, Kamila Czene⁶, Mikael Eriksson⁶, Hatef Darabi⁶, Judith S Brand⁶, Stig E Bojesen^{7,8,9}, Børge G Nordestgaard^{7,8,9}, 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^{21,22,23,24,25}, Muhammad G Kibriya^{21,22}, Alice S Whittemore^{26,27}, Esther M John^{26,27,28}, 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 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, Daehee Kang^{48,49,50}, Ji-Yeob Choi^{49,50}, Sue K Park^{48,49,50}, Keun-Young Yoo⁵¹, Keitaro Matsuo⁵², Hidemi Ito⁵³, Hiroji Iwata⁵⁴, Kazuo Tajima⁵⁵, Pascal Guénéel^{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 WR 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 WM 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^{92,93,94}, Chia-Ni Hsiung⁹³, Pei-Ei Wu^{92,93}, Ming-Feng Hou^{95,96}, Vessela N Kristensen^{97,98,99}, 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}, Petra Peeters^{104,105}, Eiliv Lund¹⁰⁶, Malin Sund¹⁰⁷, Kay-Tee Khaw¹⁰⁸, Marc J Gunter¹⁰⁵, Domenico Palli¹⁰⁹, Lotte Maxild Mortensen¹¹⁰, Laure Dossus^{111,112}, Jose-Maria Huerta¹¹³, Alfons Meindl¹¹⁴, Rita K Schmutzler^{115,116,117}, Christian Sutter¹¹⁸, Rongxi Yang^{59,60}, Kenneth Muir^{119,120}, Artitaya Lophatananon¹¹⁹, Sarah Stewart-Brown¹¹⁹, Pornthep Siriwanarangsarn¹²¹, Mikael Hartman^{122,123}, Hui Miao¹²³, Kee Seng Chia¹²³, Ching Wan Chan¹²⁴, Peter A Fasching^{125,126}, Alexander Hein¹²⁵, Matthias W Beckmann¹²⁵, Lothar Haeberle¹²⁵, Hermann Brenner^{127,128}, Aida Karina Dieffenbach^{127,128}, Volker Arndt¹²⁷, Christa Stegmaier¹²⁹, Alan Ashworth¹³, Nick Orr¹³, Minouk J Schoemaker¹¹, Anthony J Swerdlow^{11,130}, Louise Brinton³⁸, Montserrat Garcia-Closas^{11,13}, Wei Zheng⁶⁶, Sandra L Halverson⁶⁶, Martha Shrubsole⁶⁶, Jirong Long⁶⁶, Mark S Goldberg^{131,132}, France Labrèche^{133,134}, Martine Dumont^{135,136}, Robert Winqvist¹³⁷, Katri Pylkäs¹³⁷, Arja Jukkola-Vuorinen¹³⁸, Mervi Grip¹³⁹, Hiltrud Brauch^{128,140,141,142}, Ute Hamann¹⁴³, Thomas Brüning¹⁴⁴, The GENICA Network¹², Paolo Radice¹⁴⁵, Paolo Peterlongo¹⁴⁶, Siranoush Manoukian¹⁴⁷, Loris Bernard^{148,149}, Natalia V Bogdanova¹⁵⁰, Thilo Dörk¹⁵¹, Arto Mannermaa^{152,153,154},

Vesa Kataja^{155,156}, Veli-Matti Kosma^{152,153,154}, Jaana M Hartikainen^{152,153,154}, Peter Devilee¹⁵⁷, Robert AEM Tollenaar¹⁵⁸, Caroline Seynaeve¹⁵⁹, Christi J Van Asperen¹⁶⁰, Anna Jakubowska¹⁶¹, Jan Lubinski¹⁶¹, Katarzyna Jaworska¹⁶¹, Tomasz Huzarski¹⁶¹, Suleeporn Sangrajrang¹⁶², Valerie Gaborieau¹⁶³, Paul Brennan¹⁶³, James McKay¹⁶³, Susan Slager⁴⁴, Amanda E Toland¹⁶⁴, Christine B Ambrosone¹⁶⁵, Drakoulis Yannoukacos¹⁶⁶, Maria Kabisch¹⁴³, Diana Torres^{143,167}, Susan L Neuhausen¹⁶⁸, Hoda Anton-Culver¹⁶⁹, Craig Luccarini⁵, Caroline Baynes⁵, Shahana Ahmed⁵, Catherine S Healey⁵, Daniel C Tessier¹⁷⁰, Daniel Vincent¹⁷⁰, Francois Bacot¹⁷⁰, Guillermo Pita⁶², M Rosario Alonso⁶², Nuria Álvarez⁶², Daniel Herrero⁶², Jacques Simard^{135,136}, Paul PDP Pharoah^{1,5}, Peter Kraft³, Alison M Dunning⁵, Georgia Chenevix-Trench², Per Hall⁶, and Douglas F Easton^{1,5}

Affiliations

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, MA, 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 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 ²²Department of Health Studies, University of Chicago, Chicago, Illinois ²³Department of Medicine, University of Chicago, Chicago, Illinois ²⁴Department of Human Genetics, University of Chicago, Chicago, Illinois ²⁵Comprehensive Cancer Center, University of Chicago, Chicago, Illinois ²⁶Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford, CA,

USA ²⁷Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA ²⁸Department of Epidemiology, Cancer Prevention Institute of California, Fremont, CA, USA ²⁹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington ³⁰Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina ³¹Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York ³²Norway Cancer Registry, Norway ³³Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, Melbourne School of Population Health, Melbourne, Australia ³⁴Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, 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, MD, USA ³⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, 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, MN, USA ⁴⁴Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA ⁴⁵Department of Clinical Genetics, VU University Medical Center, section Oncogenetics, 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, Aichi, Japan ⁵⁴Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya, Japan ⁵⁵Epidemiology Center for Disease Control and Prevention, Mie University Hospital, Tsu, Mie, Japan ⁵⁶Inserm (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, 94807 Villejuif, France ⁵⁷University Paris-Sud, UMRS 1018, 94807 Villejuif, France ⁵⁸Université Paris Sorbonne Cité, UMR-S775 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 ⁶⁶Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA ⁶⁷Shanghai Center for Disease Control and Prevention, Changning, Shanghai, 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, ON, Canada ⁷³Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada ⁷⁴Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada ⁷⁵Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada ⁷⁶Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada ⁷⁷Laboratory Medicine Program, University Health Network, Toronto, ON, 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, UK ⁸⁰Oxford Biomedical Research Centre, University of Oxford, UK ⁸¹Surgery, School of Medicine, National University of Ireland, Galway 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, 3008 AE Rotterdam, The Netherlands ⁸⁷Department of Clinical Genetics, Erasmus University Medical Center, 3000 CA Rotterdam, The Netherlands ⁸⁸International Epidemiology Institute, Rockville, MD, USA ⁸⁹Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA ⁹⁰Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia ⁹¹Department of Pathology, The University of Melbourne, Melbourne, 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 (UiO), Oslo, Norway ⁹⁹Department of Clinical Molecular Biology (EpiGen), University of Oslo (UiO), Oslo, Norway ¹⁰⁰Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC, Australia ¹⁰¹Anatomical Pathology, The Alfred Hospital, Melbourne, VIC, Australia

¹⁰²Genomic Epidemiology Group, German Cancer Research Center (Deutsches Krebsforschungszentrum; DKFZ), Heidelberg, Germany ¹⁰³Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece; Hellenic Health Foundation, Athens, Greece ¹⁰⁴Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands ¹⁰⁵Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK ¹⁰⁶Department of Community Medicine, Faculty of Health Sciences, University of Tromsø - The Arctic University of Norway, Tromsø, Norway ¹⁰⁷Department of Surgical and Perioperative Sciences, Umea University, Sweden ¹⁰⁸School of Clinical Medicine, Cambridge Institute of Public Health, University of Cambridge, UK ¹⁰⁹Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute, ISPO, Florence, Italy ¹¹⁰Section for Epidemiology, Aarhus University, Aarhus, Denmark ¹¹¹Inserm Centre for research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health team, Villejuif, France ¹¹²Univ Paris Sud, Villejuif, France ¹¹³Department of Epidemiology, Consejería de Sanidad y Política Social, CIBER de Epidemiología y Salud Pública, Murcia, Spain ¹¹⁴Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany ¹¹⁵Center for Hereditary Breast and Ovarian Cancer, University Hospital Cologne, Cologne, Germany ¹¹⁶Center for Integrated Oncology (CIO), University Hospital Cologne, Cologne, Germany ¹¹⁷Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine, University of Cologne, Cologne, Germany ¹¹⁸Institute of Human Genetics, University Hospital Heidelberg, Heidelberg, Germany ¹¹⁹Division of Health Sciences, Warwick Medical school, Warwick University, Coventry, UK ¹²⁰Institute of Population Health, University of Manchester, Manchester, UK ¹²¹Ministry of Public Health, Nonthaburi, Thailand ¹²²Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore ¹²³Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore ¹²⁴Division of General Surgery, National University Health System ¹²⁵University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany ¹²⁶University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, Los Angeles, CA, USA ¹²⁷Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany ¹²⁸German Cancer Consortium (DKTK), Heidelberg, Germany ¹²⁹Saarland Cancer Registry, Saarbrücken, Germany ¹³⁰Division of Breast Cancer Research, Institute of Cancer Research, London, UK ¹³¹Department of Medicine, McGill University, Montreal, QC, Canada ¹³²Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, QC, Canada ¹³³Département de Santé environnementale et santé au travail, École de santé publique, Université de Montréal, Montreal, Quebec, Canada ¹³⁴Département de Médecine sociale et préventive, École de santé publique, Université de Montréal, Montreal, Quebec,

Canada ¹³⁵Centre Hospitalier Universitaire de Québec Research Center, Quebec City, Quebec, Canada ¹³⁶Laval University, Quebec City, Quebec, Canada ¹³⁷Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, NordLab Oulu/Oulu University Hospital, Oulu, Finland ¹³⁸Department of Oncology, Oulu University Hospital, University of Oulu, Oulu, Finland ¹³⁹Department of Surgery, Oulu University Hospital, University of Oulu, Oulu, Finland ¹⁴⁰Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany ¹⁴¹University of Tübingen, Tübingen, Germany ¹⁴²German Cancer Research Center (DKFZ), Heidelberg, Germany ¹⁴³Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany ¹⁴⁴Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany ¹⁴⁵Unit 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 ¹⁴⁶Fondazione Istituto FIRC (Italian Foundation for Cancer Research) di Oncologia Molecolare (IFOM), Milan, Italy ¹⁴⁷Unit 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 ¹⁴⁸Department of Experimental Oncology, Istituto Europeo di Oncologia, Milan, Italy ¹⁴⁹Cogentech Cancer Genetic Test Laboratory, Milan, Italy ¹⁵⁰Department of Radiation Oncology, Hannover Medical School, Hannover, Germany ¹⁵¹Gynaecology Research Unit, Hannover Medical School, Hannover, Germany ¹⁵²School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland ¹⁵³Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio, Finland ¹⁵⁴Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland ¹⁵⁵Cancer Center, Kuopio University Hospital, Kuopio, Finland ¹⁵⁶Central Finland Hospital District, Jyväskylä Central Hospital, Jyväskylä, Finland ¹⁵⁷Department of Human Genetics and Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands ¹⁵⁸Department of Surgical Oncology, Leiden University Medical Center, Leiden, The Netherlands ¹⁵⁹Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands ¹⁶⁰Department of Clinical Genetics, Leiden University Medical Center Leiden, The Netherlands ¹⁶¹Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland ¹⁶²National Cancer Institute, Bangkok, Thailand ¹⁶³International Agency for Research on Cancer, Lyon, France ¹⁶⁴Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH, USA ¹⁶⁵Roswell Park Cancer Institute, Buffalo, NY, USA ¹⁶⁶Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos", Aghia Paraskevi Attikis, Athens, Greece ¹⁶⁷Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota, Colombia ¹⁶⁸Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA 91010, USA ¹⁶⁹Department of Epidemiology, University of California Irvine, Irvine, CA, USA

¹⁷⁰McGill University and Génome Québec Innovation Centre, Montréal, Québec, Canada

Acknowledgments

The authors wish to 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 n° 223175 (HEALTH-F2-2009-223175) (COGS). Meetings of the BCAC have been funded by the European Union COST programme [BM0606]. Genotyping of 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 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. Combining the GWAS data was supported in part by The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148065-01 (DRIVE, part of the GAME-ON initiative). For a full description of funding and acknowledgments, see Supplementary Note.

References

1. Kamangar F, Dores GM, Anderson WF. 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*. 2006; 24:2137–50. [PubMed: 16682732]
2. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007; 447:1087–93. [PubMed: 17529967]
3. Hunter DJ, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007; 39:870–4. [PubMed: 17529973]
4. Stacey SN, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2008; 40:703–6. [PubMed: 18438407]
5. Stacey SN, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2007; 39:865–9. [PubMed: 17529974]
6. Ahmed S, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet*. 2009; 41:585–90. [PubMed: 19330027]
7. Zheng W, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet*. 2009; 41:324–8. [PubMed: 19219042]
8. 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*. 2009; 41:579–84. [PubMed: 19330030]
9. Turnbull C, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet*. 2010; 42:504–7. [PubMed: 20453838]
10. Antoniou AC, 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*. 2010; 42:885–92. [PubMed: 20852631]
11. Fletcher O, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst*. 2011; 103:425–35. [PubMed: 21263130]
12. Haiman CA, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet*. 2011; 43:1210–4. [PubMed: 22037553]
13. Ghoussaini M, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet*. 2012; 44:312–8. [PubMed: 22267197]
14. 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*. 2012; 21:5373–84. [PubMed: 22976474]
15. Long J, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet*. 2012; 8:e1002532. [PubMed: 22383897]

16. Michailidou K, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet.* 2013; 45:353–61. [PubMed: 23535729]
17. Garcia-Closas M, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet.* 2013; 45:392–8. [PubMed: 23535733]
18. Bojesen SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet.* 2013; 45:371–84. [PubMed: 23535731]
19. Milne RL, et al. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Hum Mol Genet.* 2014
20. 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.* 2014; 46:886–90. [PubMed: 25038754]
21. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 2010; 11:499–511. [PubMed: 20517342]
22. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012; 44:955–9. [PubMed: 22820512]
23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26:2190–1. [PubMed: 20616382]
24. Cox A, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet.* 2007; 39:352–8. [PubMed: 17293864]
25. Lin WY, et al. Identification and characterisation of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. *Hum Mol Genet.* 2014
26. Meijers-Heijboer H, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet.* 2002; 31:55–9. [PubMed: 11967536]
27. CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10 860 breast cancer cases 9 065 controls from 10 studies. *Am J Hum Genet.* 2004; 74:1175–82. [PubMed: 15122511]
28. Gudbjartsson DF, et al. Many sequence variants affecting diversity of adult human height. *Nat Genet.* 2008; 40:609–15. [PubMed: 18391951]
29. Kuchenbaecker KB, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet.* 2015
30. Udler MS, Tyrer J, Easton DF. Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. *Genet Epidemiol.* 2010; 34:463–8. [PubMed: 20583289]
31. Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014; 46:310–5. [PubMed: 24487276]
32. 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.* 2014; 24:1–13. [PubMed: 24196873]
33. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013; 155:934–47. [PubMed: 24119843]
34. Wang Z, et al. RNF115/BCA2 E3 ubiquitin ligase promotes breast cancer cell proliferation through targeting p21Waf1/Cip1 for ubiquitin-mediated degradation. *Neoplasia.* 2013; 15:1028–35. [PubMed: 24027428]
35. 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.* 2013; 19:253–62. [PubMed: 23821363]
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.* 2014; 23:658–69. [PubMed: 24493630]
37. Stevens KN, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. *Cancer Res.* 2012; 72:1795–803. [PubMed: 22331459]

38. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490:61–70. [PubMed: 23000897]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

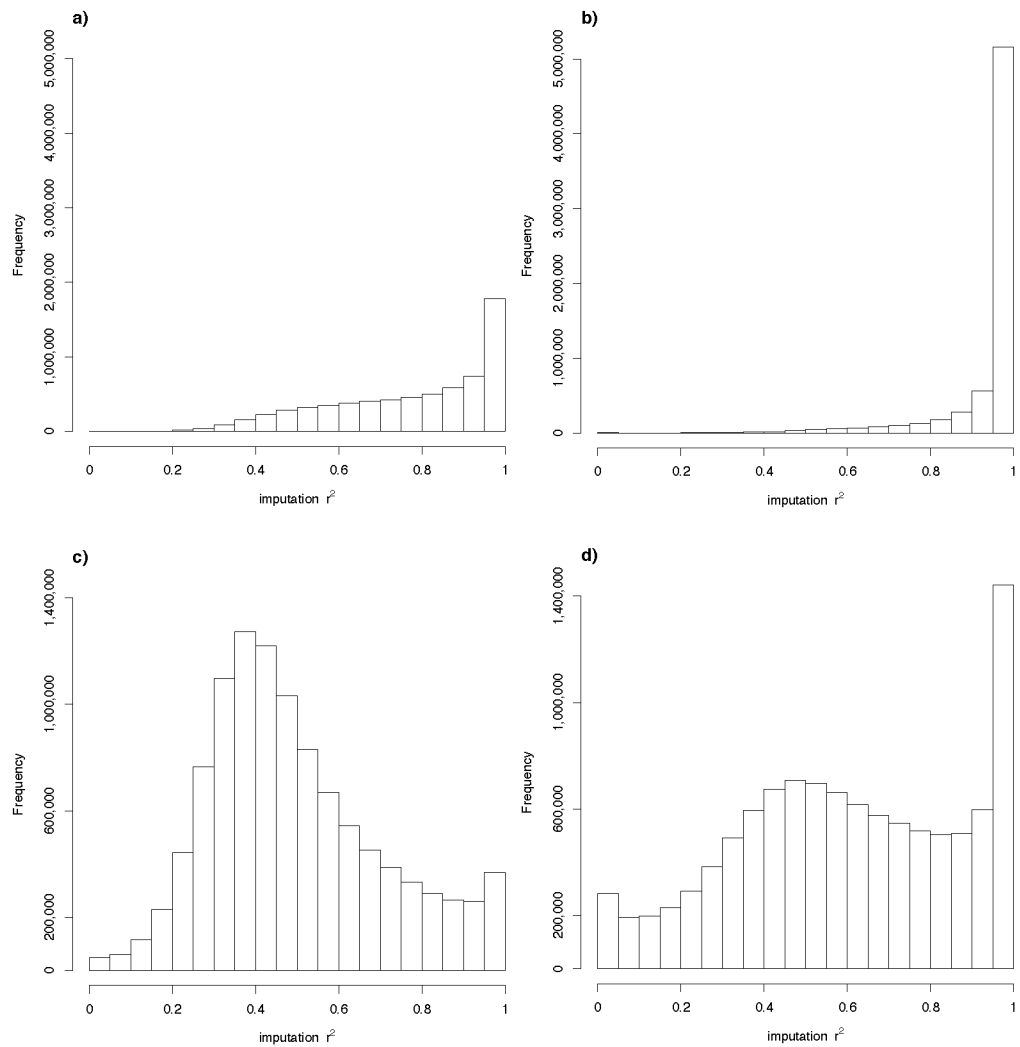


Figure 1. Histograms of the imputation r^2 **a)** Histogram of the imputation r^2 for the iCOGS for variants with $MAF > 0.05$ **b)** Histogram of the imputation r^2 for the UK2 GWAS for variants with $MAF > 0.05$ **c)** Histogram of the imputation r^2 for the iCOGS for variants with $MAF \leq 0.05$ **d)** Histogram of the imputation r^2 for the UK2 GWAS for variants with $MAF \leq 0.05$.

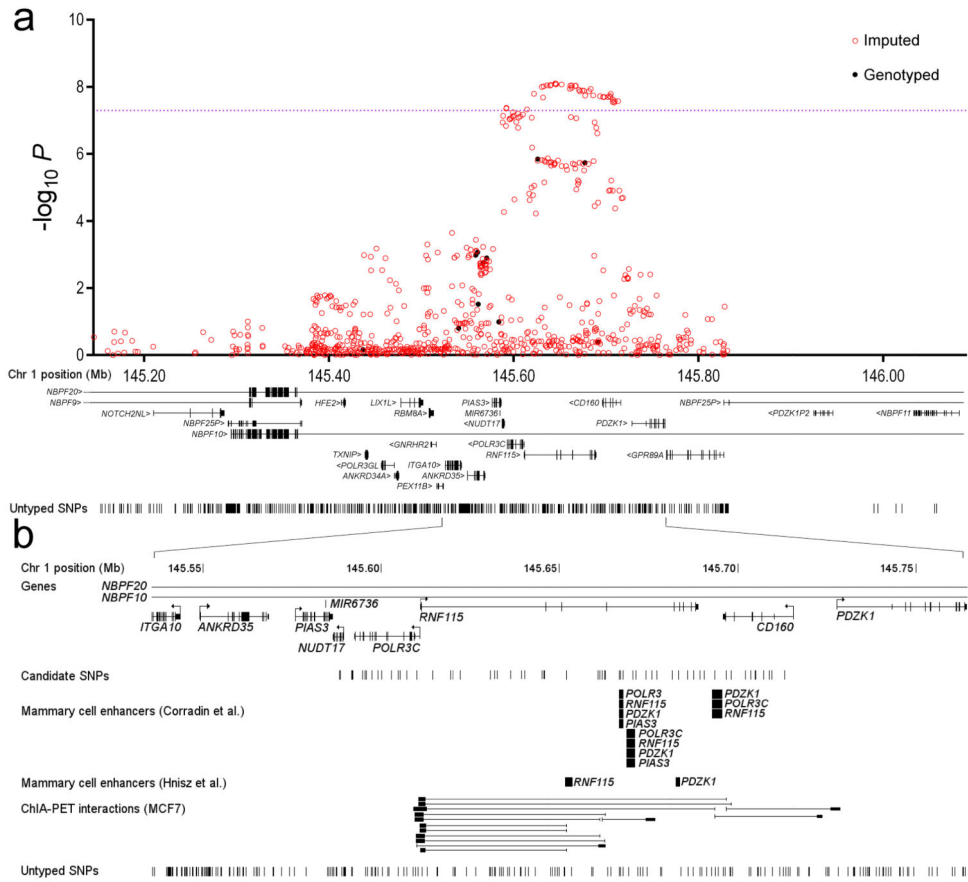


Figure 2. The chromosome 1 locus tagged by rs12405132 **a)** The Manhattan Plot displays the strength of genetic association ($-\log_{10} P$) versus chromosomal position (Mb), where each dot presents a genotyped (solid black dot) or imputed (red circle) SNP (in the iCOGS stage). The purple horizontal line represents the threshold for genome-wide significance ($P=5 \times 10^{-8}$). Gene structures are depicted as well as the location of SNPs with $MAF > 0.01$ which 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 labelled with their predicted target genes. A subset of ChIA-PET interactions in MCF7 cells (mediated by either RNAPolIII or ERa) between enhancers and their target gene promoters are also shown.

Table 1

Results for the 15 regions with combined $P < 5 \times 10^{-8}$. Results are shown for the strongest associated variant in the region.

Best variant	Locus	Position ²	Alleles ³	EAF ⁴	r ² ⁵	GWAS OR (95% CI) ⁶	GWAS P ⁷	iCOGS OR (95% CI)	iCOGS P	Combined GWAS + iCOGS P	Genes within +/-2kb	Enhancers in MCF7/HMEC	eQTLs
rs12405132	1q21.1	145644984	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, NBPF10, RNF115</i>	<i>RNF115, POLR3C, PDZK1, PIAS3</i>	-
rs12048493	1q21.2	149927034	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	242034263	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	46866866	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	16187528	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	32567732	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	81538046	A/T	0.23	0.88	0.94 (0.9–0.98)	0.00302	0.93 (0.91–0.95)	4.09×10 ⁻⁹	5.00×10 ⁻¹¹	<i>ATG10</i>	-	<i>RPS23, ATP6A1L</i>
rs9257408	6p22.1	28926220	G/C	0.38	0.92	1.05 (1–1.1)	0.0372	1.05 (1.03–1.08)	4.53×10 ⁻⁷	4.84×10 ⁻⁸	-	-	-
rs4593472	7q32.3	130667121	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>FLH3663</i>	-	-
rs13365225	8p11.23	36858483	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	117209548	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	93104072	T/C	0.26	0.73	0.94 (0.9–0.98)	0.00114	0.94 (0.92–0.96)	1.06×10 ⁻⁶	4.48×10 ⁻⁹	<i>RIN3</i>	-	-
chr17:29230520	17q11.2	29230520	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	77781725	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	42399590	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>	<i>SETBP1</i>	-

¹ Chromosome² Build 37 position³ Reference/effect allele, based on the forward strand⁴ Mean effect allele frequency over all controls⁵ Imputation r² in the iCOGS samples (calculated by the average info score from IMPUTEv2)⁶ Per allele odds ratio for the minor allele relative to the major allele⁷ P value for the 1df trend test