

# Genome-wide association studies identify four ER negative–specific breast cancer risk loci

Estrogen receptor (ER)-negative tumors represent 20–30% of all breast cancers, with a higher proportion occurring in younger women and women of African ancestry<sup>1</sup>. The etiology<sup>2</sup> and clinical behavior<sup>3</sup> of ER-negative tumors are different from those of tumors expressing ER (ER positive), including differences in genetic predisposition<sup>4</sup>. To identify susceptibility loci specific to ER-negative disease, we combined in a meta-analysis 3 genome-wide association studies of 4,193 ER-negative breast cancer cases and 35,194 controls with a series of 40 follow-up studies (6,514 cases and 41,455 controls), genotyped using a custom Illumina array, iCOGS, developed by the Collaborative Oncological Gene-environment Study (COGS). SNPs at four loci, 1q32.1 (*MDM4*,  $P = 2.1 \times 10^{-12}$  and *LGR6*,  $P = 1.4 \times 10^{-8}$ ), 2p24.1 ( $P = 4.6 \times 10^{-8}$ ) and 16q12.2 (*FTO*,  $P = 4.0 \times 10^{-8}$ ), were associated with ER-negative but not ER-positive breast cancer ( $P > 0.05$ ). These findings provide further evidence for distinct etiological pathways associated with invasive ER-positive and ER-negative breast cancers.

ER-negative tumors are associated with a worse short-term prognosis<sup>3</sup> and have weaker associations with reproductive risk factors<sup>2</sup> than ER-positive tumors. There are also important differences in genetic susceptibility to these two types of tumors. *BRCA1* mutations predispose primarily to ER-negative disease, whereas most known common susceptibility loci for breast cancer show stronger associations with ER-positive than with ER-negative tumors<sup>4</sup>. Exceptions are three loci tagged by rs10069690 on chromosome 5p15 (ref. 5) (*TERT-CLPTMIL*), rs8170 at 19p13 (ref. 6) (*BABAM1*, also known as *MERIT40*) and rs2284378 at 20q11 (ref. 7), which predispose primarily to ER-negative tumors, and loci at 6q25 (ref. 8) that confer higher risk for ER-negative than for ER-positive tumors. With the aim of identifying susceptibility loci specific for invasive ER-negative disease, we analyzed three genome-wide association studies (GWAS) in populations of European ancestry and followed-up promising signals from each GWAS in the Breast Cancer Association Consortium (BCAC).

The 3 GWAS included a total of 4,193 ER-negative breast cancer cases and 35,194 controls of European ancestry drawn from 23 studies participating in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3), the Triple-Negative Breast Cancer Consortium (TNBCC) and the Combined BCAC ER-negative GWAS (C-BCAC) (Online Methods and **Supplementary Table 1**). We selected 13,276 SNPs on the basis of rank  $P$  values from the 3 GWAS, and these were genotyped in an independent set of 6,514

ER-negative cases and 41,455 controls of European ancestry from 40 BCAC studies forming part of the COGS Project (Online Methods and **Supplementary Table 1**). Samples were genotyped using the iCOGS custom Illumina Infinium array that included a total of 211,155 SNPs selected in collaboration with other cancer consortia (Online Methods). We performed a fixed-effects meta-analysis of odds ratio (OR) estimates from the GWAS and follow-up studies (quantile-quantile plot shown in **Supplementary Fig. 1**) and identified four loci newly associated with ER-negative disease at  $P < 5 \times 10^{-8}$  (**Fig. 1** and **Table 1**; cluster plots shown in **Supplementary Fig. 2**).

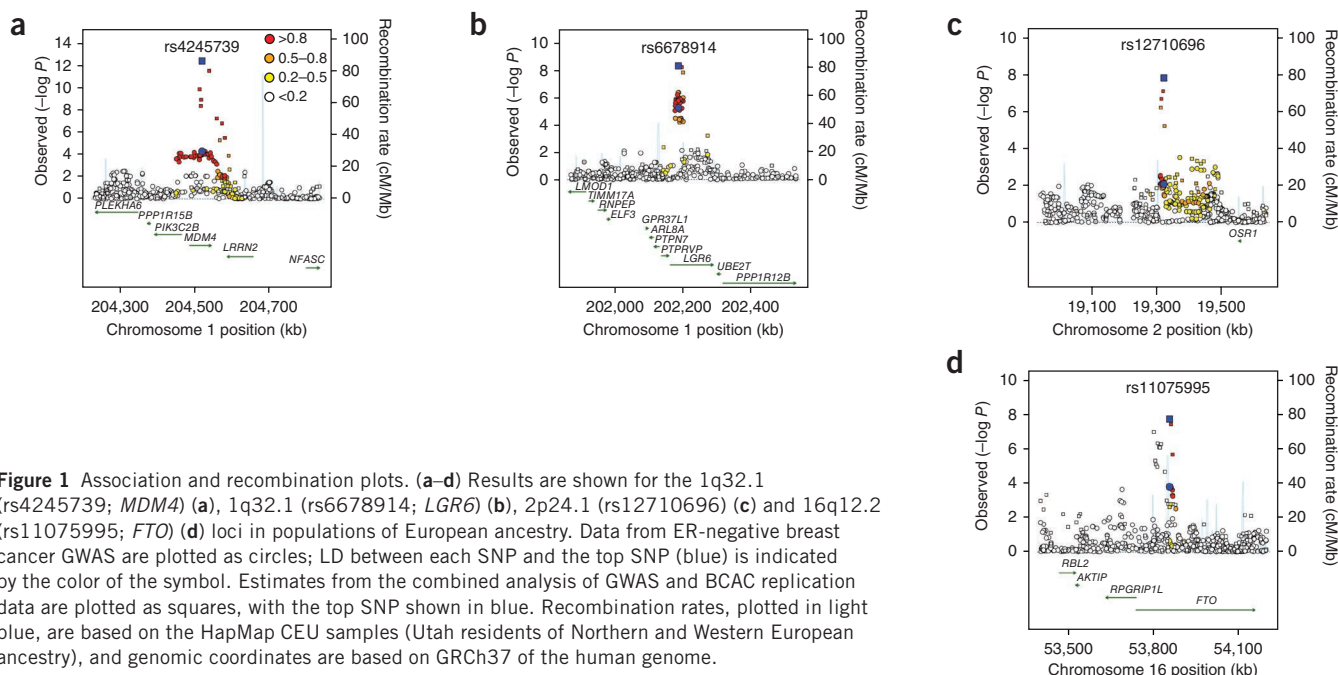
Two independently associated loci were located on chromosome 1q32.1 and were tagged by two uncorrelated ( $r^2 < 0.001$ ) markers (from reference sequence NCBI Build 36): rs4245739 ( $P = 2.1 \times 10^{-12}$ , OR = 1.14, 95% confidence interval (CI) = 1.10–1.18) and rs6678914 ( $P = 1.4 \times 10^{-8}$ , OR = 1.10, 95% CI = 1.06–1.13). Conditional analyses of the two SNPs in BCAC follow-up data showed comparable estimates, indicating that these are two distinct signals (**Supplementary Table 2**). The other two loci were located at 2p24.1 (rs12710696,  $P = 4.6 \times 10^{-8}$ , OR = 1.10, 95% CI = 1.06–1.13) and 16q12.2 (rs11075995,  $P = 4.0 \times 10^{-8}$ , OR = 1.11, 95% CI = 1.07–1.15). For each region, there was little evidence for heterogeneity of effect by study (**Table 1** and **Supplementary Fig. 3a–d**), and genotype-specific risks for rs4245739, rs6678914 and rs12710696 were consistent with a log-additive model. For rs11075995, departure from a log-additive model was significant ( $P = 0.039$ ), and genotype-specific estimates suggested a recessive effect (**Supplementary Table 3**).

The strength of the association for each SNP differed significantly by ER status, and none of the SNPs showed significant associations with ER-positive disease in the analysis of 25,227 ER-positive cases and 41,455 controls of European ancestry in BCAC (**Supplementary Tables 4** and **5**). Notably, we observed no significant differences in ORs for ER-negative tumors with and without the triple-negative phenotype (defined as ER-negative, progesterone receptor (PR)-negative and HER2-negative) for rs6678914 (1q32.1, *LGR6*), rs12710696 (2p24.1) and rs11075995 (16q12.2). However, rs4245739 (1q32.1, *MDM4*) seemed to be specific to triple-negative tumors (case-only heterogeneity  $P$  value ( $P_{\text{het}}$ ) by triple-negative status = 0.005; **Supplementary Table 5**).

None of the four SNPs showed significant ( $P < 0.05$ ) associations in studies of Asian ancestry in BCAC, and only the 16q12.2 (*FTO*) variant was associated at  $P = 0.05$  in combined analyses of studies of African-American ancestry in BCAC and the African-American Breast Cancer Consortium<sup>5</sup> (AABC; **Supplementary Table 6**). However, estimates

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

Received 14 May 2012; accepted 29 January 2013; published online 27 March 2013; doi:10.1038/ng.2561



**Figure 1** Association and recombination plots. (a–d) Results are shown for the 1q32.1 (rs4245739; *MDM4*) (a), 1q32.1 (rs6678914; *LGR6*) (b), 2p24.1 (rs12710696) (c) and 16q12.2 (rs11075995; *FTO*) (d) loci in populations of European ancestry. Data from ER-negative breast cancer GWAS are plotted as circles; LD between each SNP and the top SNP (blue) is indicated by the color of the symbol. Estimates from the combined analysis of GWAS and BCAC replication data are plotted as squares, with the top SNP shown in blue. Recombination rates, plotted in light blue, are based on the HapMap CEU samples (Utah residents of Northern and Western European ancestry), and genomic coordinates are based on GRCh37 of the human genome.

for Asian and African-American populations were not significantly different from those in Europeans ( $P > 0.05$ ), and larger studies in these populations are needed to determine whether risk associations exist. None of the markers were significantly associated with increasing age at the onset of ER-negative disease in the BCAC follow-up data ( $P_{\text{trend}} \geq 0.314$ ), although there were some differences in age-specific estimates (Supplementary Table 7). Furthermore, OR estimates were not significantly different for women with and without a family history of any breast cancer in at least one first-degree relative, and risk alleles were not over-represented in cases with a positive family history (Supplementary Table 8).

rs4245739 (1q32.1) is located in the 3' region of the *MDM4* oncogene. *MDM4* is a repressor of *TP53* and *TP73* transcription and is important for cell cycle regulation and apoptosis. rs4245739 resides in a linkage disequilibrium (LD) block of approximately 230 kb (Supplementary Fig. 4a) that also contains the tRNA<sup>Lys</sup> transcript and the genes *PIK3C2B* and *LRRN2* (Supplementary Fig. 5a). *MDM4*, tRNA<sup>Lys</sup> and *PIK3C2B* but not *LRRN2* are expressed in normal breast epithelium, breast cancer cell lines and breast tumors<sup>9–11</sup>. There are no nonsynonymous SNPs correlated with rs4245739 in the 1000 Genomes Project populations of European ancestry ( $r^2 > 0.10$ ); however, correlated SNPs are located in the promoter region of *PIK3C2B* (rs3014606,  $r^2 = 0.94$  and rs2926534,  $r^2 = 0.94$ ) and in the tRNA<sup>Lys</sup> transcript (rs11240753,  $r^2 = 0.78$  and rs4951389,  $r^2 = 0.78$ ). Variants in the *MDM4* locus correlated with rs4245739 have also been associated with breast cancer in *BRCA1* mutation carriers who have predominantly ER-negative tumors<sup>12</sup>. Thus, this region seems to be specifically associated with ER-negative disease and not with overall breast cancer risk, as suggested by a previous, smaller candidate gene study<sup>13</sup>. To our knowledge, no studies before the COGS collaboration have evaluated rs4245739 in relation to the risk of ER-negative disease.

rs6678914 on chromosome 1q32.1 is located in intron 1 of the *LGR6* gene (Supplementary Fig. 4b). *LGR6* and several other genes in this region, including *UBE2T* and *PTPN7*, are expressed in breast tumors<sup>9</sup>. A correlated SNP (rs12032424,  $r^2 = 0.96$ ) is located in a putative enhancer region in the same intron of *LGR6* in normal breast

epithelial cells, although not in the triple-negative breast cancer cell line MDA-MB-231 (Supplementary Fig. 5b). The rs6678914 SNP is not correlated with nonsynonymous SNPs in *LGR6* ( $r^2 > 0.10$  in 1000 Genomes Project populations of European ancestry).

The SNP rs12710696 on chromosome 2p24.1 is located in an intergenic region, more than 200 kb from the nearest gene (*OSR1*) (Supplementary Fig. 4c). It is possible that the allele marked by rs12710696 could influence a set of active enhancers, as the region contains multiple overlapping chromatin marks in normal breast epithelial cells and the MDA-MB-231 triple-negative breast cancer cell line (Supplementary Fig. 5c).

The signal found on chromosome 16q12.2 is located in the fat mass- and obesity-associated gene *FTO* (Supplementary Fig. 4d). This signal is tagged by rs11075995, located in a ~40-kb LD block in intron 1 of *FTO*, within an enhancer region that appears to be active in both normal and triple-negative breast cancer cells (Supplementary Fig. 5d). rs11075995 is located ~40 kb distal to a region in intron 1 that contains multiple SNPs associated with obesity in the Genetic Investigation of ANthropometric Traits (GIANT) Consortium<sup>14,15</sup>, as well as a SNP associated with overall breast cancer risk (rs17817449)<sup>8</sup>. rs11075995 is not correlated with any of the previously reported SNPs associated with obesity at genome-wide significant levels in GIANT or with rs17817449 ( $P = 3.7 \times 10^{-60}$ , based on 123,864 subjects in GIANT; ref. 15). However, rs11075995 is associated with body mass index (BMI), both in GIANT ( $P = 1.51 \times 10^{-6}$ , based on 121,427 subjects) and our control population ( $P = 2.8 \times 10^{-5}$ , based on 20,952 controls in iCOGS; data not shown). Analyses adjusting and stratifying by BMI on the basis of 3,071 ER-negative cases and 20,130 controls from 19 studies genotyped on the iCOGS array indicated that the association between rs11075995 and ER-negative disease is not explained or modified by our measure of BMI (BMI-adjusted OR = 1.16, 95% CI = 1.09–1.24,  $P = 1.1 \times 10^{-5}$ ;  $P_{\text{interaction}} = 0.912$ ; data not shown). Furthermore, conditional analyses indicated that the ER-negative disease-specific signal (rs11075995) is independent of rs17817449 (Supplementary Table 2). This finding adds to the increasing evidence of distinct signals at the same locus for different subtypes of cancers occurring at the same site, including, for example, 5p15.33 (*TERT-CLPTM1L*)<sup>16</sup> and 14q24.1

(*RAD51B*, also known as *RAD51L1*)<sup>8</sup> in breast cancer and 5p15.33 (*TERT-CLPTMIL*)<sup>16</sup> and *HNF1B*<sup>17</sup> in ovarian cancer. Detailed fine mapping of known and newly identified breast cancer-associated regions will be required to determine whether additional subtype-specific signals exist in these regions.

In an attempt to investigate the likely genes responsible in the observed risk associations, we examined associations between SNPs

with available genotype (rs4245739, rs12710696 and rs6678914) and RNA expression in data from 382 primary breast tumors, including 81 ER-negative samples in The Cancer Genome Atlas (TCGA) database. None of the associations were significant after Bonferroni adjustment for multiple comparisons, whether considering only the immediately neighboring genes or all genes within a 1-Mb window of the lead SNP (data not shown).

**Table 1 Associations of SNPs and ER-negative breast cancer risk in populations of European ancestry**

SNP	Cytoband	Gene	Position <sup>a</sup>	Stage	T/I <sup>b</sup>	Studies	Cases	Controls	RAF	OR (95% CI)	P	<i>P</i> <sub>het</sub>	<i>I</i> <sup>2</sup> study het. (%) <sup>c</sup>
rs4245739	1q32.1	<i>MDM4</i>	202785465	<b>GWAS</b>									
				BPC3	I	7	2,069	25,385	0.27	1.07 (0.97–1.17)	0.177		
				TNBCC	I	11	1,562	3,399		1.20 (1.08–1.32)	4.6 × 10 <sup>-4</sup>		
				C-BCAC	I	5	562	6,410	0.28	1.17 (1.02–1.35)	0.024		
				<b>Follow-up</b>									
				BCAC/iCOGS	T	40	6,512	41,451	0.26	1.13 (1.09–1.18)	8.5 × 10 <sup>-9</sup>		
<b>Meta-analysis</b>						<b>63</b>	<b>10,705</b>	<b>76,645</b>	<b>0.26</b>	<b>1.14</b> <b>(1.10–1.18)</b>	<b>2.1 × 10<sup>-12</sup></b>	<b>0.413</b>	<b>3.2</b>
rs6678914	1q32.1	<i>LGR6</i>	20045399	<b>GWAS</b>									
				BPC3	I/T	7	2,069	25,385	0.59	1.12 (1.03–1.22)	0.007		
				TNBCC	T	11	1,562	3,399	0.59	1.16 (1.05–1.27)	0.003		
				C-BCAC	I/T	5	562	6,410	0.59	1.15 (1.01–1.30)	0.032		
				<b>Follow-up</b>									
				BCAC/iCOGS	T	40	6,514	41,452	0.59	1.08 (1.04–1.12)	1.8 × 10 <sup>-4</sup>		
<b>Meta-analysis</b>						<b>63</b>	<b>10,707</b>	<b>76,646</b>	<b>0.59</b>	<b>1.10</b> <b>(1.06–1.13)</b>	<b>1.4 × 10<sup>-8</sup></b>	<b>0.481</b>	<b>0.0</b>
rs12710696	2p24.1	Non-genic	19184284	<b>GWAS</b>									
				BPC3	I	7	2,069	25,385	0.37	1.05 (0.96–1.14)	0.304		
				TNBCC	I	11	1,562	3,399		1.17 (1.06–1.29)	0.001		
				C-BCAC	I	5	562	6,410	0.37	1.00 (0.88–1.14)	0.947		
				<b>Follow-up</b>									
				BCAC/iCOGS	T	40	6,512	41,453	0.36	1.10 (1.06–1.15)	1.4 × 10 <sup>-6</sup>		
<b>Meta-analysis</b>						<b>63</b>	<b>10,705</b>	<b>76,647</b>	<b>0.36</b>	<b>1.10</b> <b>(1.06–1.13)</b>	<b>4.6 × 10<sup>-8</sup></b>	<b>0.464</b>	<b>0.0</b>
rs11075995	16q12.2	<i>KIAA1752-FTO</i>	52412792	<b>GWAS</b>									
				BPC3	I	7	2,069	25,385	0.24	1.15 (1.04–1.28)	0.008		
				TNBCC	I	11	1,562	3,399		1.15 (1.03–1.28)	0.010		
				C-BCAC	I	5	562	6,410	0.24	1.09 (0.92–1.28)	0.328		
				<b>Follow-up</b>									
				BCAC/iCOGS	T	40	6,513	41,453	0.24	1.10 (1.05–1.15)	4.2 × 10 <sup>-5</sup>		
<b>Meta-analysis</b>						<b>63</b>	<b>10,706</b>	<b>76,647</b>	<b>0.24</b>	<b>1.11</b> <b>(1.07–1.15)</b>	<b>4.0 × 10<sup>-8</sup></b>	<b>0.079</b>	<b>24.3</b>

Results are shown for the SNPs showing the strongest association in four loci reaching association  $P < 5 \times 10^{-8}$  in meta-analyses of GWAS and follow-up data. RAF, risk allele frequency; freq., frequency.

<sup>a</sup>NCBI Build 36. <sup>b</sup>Imputed (I) and typed (T) SNPs: rs6678914 was typed in one BPC3 study (WGHS), three C-BCAC studies (ABCFS, SASBCAC, UK2) and all TNBC studies and imputed in all other GWAS studies. <sup>c</sup>Result of *Q* test for heterogeneity of estimated ORs.

To provide a comprehensive analysis of common genetic loci for ER-negative breast cancer, we also evaluated associations between 67 known loci for overall breast cancer risk (26 previously reported and 41 newly identified<sup>8</sup>) and ER-negative disease. On the basis of our meta-analysis of 10,707 ER-negative cases and 76,649 controls, 7 regions influenced risk of ER-negative disease at  $P < 5 \times 10^{-8}$ : 1p36.22 (*PEX14*), 5p15 (*TERT-CLPTM1L*), 2 independent loci at 6q25.1 (*ESR1*), 12p11.22 (*PTHLH*), 16q12.1 (*TOX3*) and 19p13.1 (*BABAM1*) (**Supplementary Table 9**). Only seven loci identified so far, the four reported here and the three previously reported located at 5p15 (ref. 5), 19p13.1 (ref. 6) and 20q11 (ref. 7), are specific to ER-negative disease.

In summary, our analyses provide further evidence for distinct etiological pathways for invasive ER-positive and ER-negative breast cancers. Fine mapping and functional studies of the susceptibility loci for ER-negative disease should provide important insights into the biological mechanisms of ER-negative breast cancer, potentially leading to the identification of new targets for therapy and prevention of this aggressive form of breast cancer.

**URLs.** BCAC, <http://www.srl.cam.ac.uk/consortia/bcac/index.html>; CIMBA, <http://www.srl.cam.ac.uk/consortia/cimba/index.html>; COGS, <http://www.cogseu.org/>; GIANT, [http://www.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium](http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium); OCAC, <http://www.srl.cam.ac.uk/consortia/ocac/index.html>; PRACTICAL, <http://www.srl.cam.ac.uk/consortia/practical/index.html>; TCGA, <http://www.cancergenome.nih.gov/>; 1000 Genomes Project, <http://www.1000genomes.org/>; GLU (Genotype Library and Utilities), <http://code.google.com/p/glu-genetics/>; UCSC Genome Browser, <http://genome.ucsc.edu/>.

## METHODS

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

**Accession codes.** Reference sequences for the human genome of the regions containing the following genes are available at NCBI under the indicated accessions: *LRRN2*, [NC\\_000001.10](#); *UBE2T*, [NC\\_000001.10](#); *PTPN7*, [NC\\_000001.10](#); *PEX14*, [NC\\_000001.10](#); *LGR6*, [NC\\_000001.10](#); *MDM4*, [NC\\_000001.10](#); *TP73*, [NC\\_000001.10](#); *PIK3C2B*, [NC\\_000001.10](#); *OSR1*, [NC\\_000002.11](#); *TERT*, [NC\\_000005.9](#); *CLPTM1L*, [NC\\_000005.9](#); *ESR1*, [NC\\_000006.11](#); *PTHLH*, [NC\\_000012.11](#); *RAD51B*, [NC\\_000014.8](#); *FTO*, [NC\\_000016.9](#); *TOX3*, [NC\\_000016.9](#); *HNF1B*, [NC\\_000017.10](#); *BRCA1*, [NC\\_000017.10](#); *TP53*, [NC\\_000017.10](#); *MERIT40*, [NC\\_000019.9](#).

Note: Supplementary information is available in the [online version of the paper](#).

## ACKNOWLEDGMENTS

The authors wish to thank all the individuals who took part in these studies and all the researchers, clinicians and administrative staff who have enabled this work to be carried out. We are very grateful to Illumina, in particular J. Stone, S. McBean, J. Hadlington, A. Mustafa and K. Cook, for their help with designing the array. BCAC is funded by Cancer Research UK (C1287/A10118 and C1287/A12014) and by the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS). Meetings of BCAC have been funded by the European Union European Cooperation in Science and Technology (COST) programme (BM0606). BPC3 is funded by US National Cancer Institute cooperative agreements U01-CA98233, U01-CA98710, U01-CA98216 and U01-CA98758 and the Intramural Research Program of the US National Institutes of Health (NIH)/National Cancer Institute, Division of Cancer Epidemiology and Genetics. TNBCC is supported by Mayo Clinic Breast Cancer Study (MCBCS) (US NIH grants CA122340 and a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201)), grants from the Komen Foundation for the Cure and the Breast Cancer Research Foundation. Genotyping

on the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710), US NIH grant CA122340, the Komen Foundation for the Cure, the Breast Cancer Research Foundation, the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer program (J. Simiard and D.E.) and Ministry of Economic Development, Innovation and Export Trade of Quebec grant PSR-SIIRI-701 (J. Simiard, D.E. and P.H.). J. Simiard holds the Canada Research Chair in Oncogenetics. Combination of the GWAS data was supported in part by US NIH Cancer Post-Cancer GWAS initiative grant U19 CA 148065-01 (DRIVE, part of the GAME-ON initiative) and Breakthrough Breast Cancer Research.

## AUTHOR CONTRIBUTIONS

M.G.-C., F.J.C., S.L., K. Michailidou, M.K.S., P.D.P.P., C.V., D.F.E., C.A.H. and P. Kraft formed the writing group and drafted the manuscript. M.G.-C. coordinated the writing group. F.J.C., S.L., K. Michailidou, D.F.E., C.A.H. and P. Kraft performed statistical analyses of GWAS data. M.G.-C. and M.N.B. performed statistical analyses of BCAC follow-up studies and meta-analyses. P. Kraft coordinated the BPC3 GWAS, and M.G.-C., E.R., H.S.F., L.L.M., J.E.B., W.C.W., D.J.H. and S.J.C. led individual studies in the BPC3 scan. F.J.C. and C.V. coordinated the TNBCC GWAS, and D.E., P. Miron, P.A.F., J.C.-C., J.C., A.A., H.N., H. Brauch and G.G.G. led individual studies in the TNBCC scan. D.E. coordinated the C-BCAC GWAS, and H.N., J.L.H., J.C.-C. and P.H. led individual studies in the C-BCAC scan. D.F.E. conceived and coordinated the synthesis of the iCOGS array and led BCAC. P.H. coordinated COGS, and J.B. led the BCAC genotyping working group. A.G.-N., G.P., M.R.A., D.V., F.B., D.C.T. and F.J.C. coordinated genotyping of the iCOGS array. M.G.-C., P.D.P.P. and M.K.S. led the pathology working group in BCAC. M.E.S. was the lead pathologist in BCAC. W.J.H. performed automated scoring of tissue microarrays. A.M.D. and G.C.-T. led the quality control working group. J.D. and N.O. provided bioinformatics support. S.K.R. and G.A.C. performed FunciSNP bioinformatics analyses. M.K.B. and Q. Wang provided data management support for BCAC. G.G., A.A., A. Broeks, A.B.E., A.C., U.H., A.-S.D., A.G.U., A.H., A.H.W., A.I., the ABCTB Investigators, A.J.-V., A.J., A.K.G., R.W., A. Lindblom, A. Lophatananon, A.M.D., A.M.M., A.M.W.v.d.O., A.R., A. Swerdlow, A. Schneeweiss, B.B., B.E.H., B.G.N., B.M.-M., B.P., C.B., C.B.A., C.-Y.C., C.C., C.D.B., C.-N.H., C.H.M.v.D., C.H.Y., C.J., C.M., C.M.S., C.O., C.R., C.-Y.S., C. Sohn, C. Stegmaier, C.-C.T., C.T., C.W.C., D.C., D.C.T., D.F.-J., D.G., D.I.C., D.J.P., D.J.S., D.K., D.L., D.O.S., D.S., D.T., D.V.D.B., E.D., C.V., E.J.R., E.M.J., E.M., E.V.B., E.W., F.A., FBSC, F.C.-C., F.C., F.H., F.L., F.M., F.R., F.S., G.A.C., G.C.-T., G.K.C., G.S., G.W.M., H.A.-C., H.C., H.F., H. Ito, H. Iwata, H. Müller, H. Miao, H.M.-H., H.P., H.T., H.W., I.d.S.S., I.K., I.L.A., I.T., J.A.K., J.D.F., J.E.O., J.I.A.P., J.J.H., J. Long, J. Lubinski, J. Liu, J. Lissowska, J.L.R.-G., J.M.H., J.P., J. Stone, J. Simard, J.W., J.-C.Y., K. Aittomäki, K. Aaltonen, K.C., K.D., K.J., K.-T.K., K.L., K. Muir, K. Matsuo, K.P., K.S., K.S.C., L. Bernard, L. Baglietto, L. Bernstein, L. Beckmann, L.D., L.G., L.J.V.V., L.N.K., L.S., M.B., M.C.S., M.D., M.F.P., M.G.S., M. Jones, M. Johansson, M.J.H., M.J.K., M.K., M.K.B., M.L., M.M.G., M.P.L., M. Shrubsole, M. Shah, M.W.B., M.W.R.R., N.A.M.T., N.D., N.G.M., N.J., N.M., N.N.A., N.R., N.S., N.V.B., O.F., P.G., P.H., P.H.P., P. Kerbrat, P.L.-P., P.L., P. Menéndez, P.N., P.P., P.R., P. Siriwanarangsang, P. Sharma, P.-E.W., Q.C., Q. Wang, Q. Waisfisz, R.B., R.G.Z., R.H., R.K., R.K.S., R.L.M., R.M.M., R.N.H., R.P., R.A.E.M.T., R. Tumino, R. Travis, S.A.I., S.E.B., S.E.H., S.F.N., S.G., S.H.T., S.K., S.K.R., S.L.D.-H., S.M., S.M.J., S. Nickels, S. Nyante, S.P.B., S. Sangrajrang, S.S.-B., S. Slager, S.S.C., T.A.M., T.B., T.D., T.H., T.T., V.A., V. Kristensen, V. Kataja, V.-M.K., W.B., W.L., W.R.D., W.T., X.-O.S., X.W., Y.F., Y.-T.G., Y.-D.K. and Y.Y. contributed to GWAS and/or BCAC follow-up studies. M.G.-C., F.J.C., S.L., K. Michailidou, M.K.S., P.D.P.P., C.V., D.F.E., C.A.H., P. Kraft, M.N.B., E.R., H.S.F., L.L.M., J.E.B., W.C.W., D.J.H., S.J.C., D.E., P.A.F., J.C.-C., J.C., A. Broeks, H.N., H. Brauch, H. Brenner, G.P., G.G.G., J.L.H., P. Miron, J.B., A.G.-N., M.R.A., D.V., F.B., M.E.S., W.J.H., G.G., A.A., A. Beck, A.B.E., A.C., U.H., A.-S.D., A.G.U., A.H., A.H.W., A.I., the ABCTB Investigators, A.J.-V., A.J., A.K.G., R.W., A. Lindblom, A. Lophatananon, A.M.D., A.M.M., A.M.W.v.d.O., A.R., A. Swerdlow, A. Schneeweiss, B.B., B.E.H., B.G.N., B.M.-M., B.P., C.B., C.B.A., C.-Y.C., C.C., C.D.B., C.-N.H., C.H.M.v.D., C.H.Y., C.J., C.M., C.M.S., C.O., C.R., C.-Y.S., C. Sohn, C. Stegmaier, C.-C.T., C.T., C.W.C., D.C., D.C.T., D.F.-J., D.G., D.I.C., D.J.P., D.J.S., D.K., D.L., D.O.S., D.S., D.T., D.V.D.B., E.D., C.V., E.J.R., E.M.J., E.M., E.V.B., E.W., F.A., FBSC, F.C.-C., F.C., F.H., F.L., F.M., F.R., F.S., G.A.C., G.C.-T., G.K.C., G.S., G.W.M., H.A.-C., H.C., H.F., H. Ito, H. Iwata, H. Müller, H. Miao, H.M.-H., H.P., H.T., H.W., I.d.S.S., I.K., I.L.A., I.T., J.A.K., J.D.F., J.E.O., J.I.A.P., J.J.H., J. Long, J. Lubinski, J. Liu, J. Lissowska, J.L.R.-G., J.M.H., J.P., J. Stone, J. Simard, J.W., J.-C.Y., K. Aittomäki, K. Aaltonen, K.C., K.D., K.J., K.-T.K., K.L., K. Muir, K. Matsuo, K.P., K.S., K.S.C., L. Bernard, L. Baglietto, L. Bernstein, L. Beckmann, L.D., L.G., L.J.V.V., L.N.K., L.S., M.B., M.C.S., M.D., M.F.P., M.G.S., M. Jones, M. Johansson, M.J.H., M.J.K., M.K., M.K.B., M.L., M.M.G., M.P.L., M. Shrubsole, M. Shah, M.W.B., M.W.R.R., N.A.M.T., N.D., N.G.M., N.J., N.M., N.N.A., N.O., N.R., N.S., N.V.B., O.F., P.G., P.H., P.H.P., P. Kerbrat, P.L.-P., P.L.,



P. Menénde, P.N., P.P., P.R., P. Siriwanarangsana, P. Sharma, P.-E.W., Q.C., Q. Wang, Q. Waisfisz, R.B., R.G.Z., R.H., R.K., R.K.S., R.L.M., R.M.M., R.N.H., R.P., R.A.E.M.T., R. Tumino, R. Travis, S.A.I., S.E.B., S.E.H., S.F.N., S.G., S.H.T., S.K., S.K.R., S.L.D.-H., S.M., S.M.J., S. Nickels, S. Nyante, S.P.B., S. Sangrajrang, S.S.-B., S. Slager, S.S.C., T.A.M., T.B., T.D., T.H., T.T., V.A., V. Kristensen, V. Kataja, V.-M.K., W.B., W.L., W.R.D., W.T., X.-O.S., X.W., Y.F., Y.-T.G., Y.-D.K. A. Mannermaa, A. Meindl, W.Z., P.D., M.S.G. and Y.Y. 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>.

- Chu, K.C. & Anderson, W.F. Rates for breast cancer characteristics by estrogen and progesterone receptor status in the major racial/ethnic groups. *Breast Cancer Res. Treat.* **74**, 199–211 (2002).
- Yang, X.R. *et al.* Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J. Natl. Cancer Inst.* **103**, 250–263 (2011).
- Blows, F.M. *et al.* Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med.* **7**, e1000279 (2010).
- Mavaddat, N., Antoniou, A.C., Easton, D.F. & Garcia-Closas, M. Genetic susceptibility to breast cancer. *Mol. Oncol.* **4**, 174–191 (2010).
- 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).
- Stevens, K.N. *et al.* 19p13.1 is a triple negative-specific breast cancer susceptibility locus. *Cancer Res.* **72**, 1795–1803 (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).
- Michailidou, K. *et al.* Large-scale genotyping identifies 41 new breast cancer susceptibility loci. *Nat. Genet.* published online; doi:10.1038/ng.2563 (27 March 2013).
- Turashvili, G. *et al.* Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer* **7**, 55 (2007).
- Graham, K. *et al.* Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br. J. Cancer* **102**, 1284–1293 (2010).
- Wang, H. & Yan, C. A small-molecule p53 activator induces apoptosis through inhibiting MDMX expression in breast cancer cells. *Neoplasia* **13**, 611–619 (2011).
- Couch, F.J. *et al.* Genome-wide association study in *BRCA1* mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet.* **9**, e1003212 (2013).
- Atwal, G.S. *et al.* Altered tumor formation and evolutionary selection of genetic variants in the human *MDM4* oncogene. *Proc. Natl. Acad. Sci. USA* **106**, 10236–10241 (2009).
- Frayling, T.M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
- Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
- Bojesen, S.E. *et al.* Multiple independent *TERT* variants associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* published online; doi:10.1038/ng.2566 (27 March 2013).
- Shen, H. *et al.* Epigenetic analysis leads to identification of *HNF1B* as a subtype-specific susceptibility gene for ovarian cancer. *Nat. Comm.* published online; doi:10.1038/ncomms2629 (27 March 2013).

Montserrat Garcia-Closas<sup>1,2,180</sup>, Fergus J Couch<sup>3,180</sup>, Sara Lindstrom<sup>4,180</sup>, Kyriaki Michailidou<sup>5,180</sup>, Marjanka K Schmidt<sup>6,180</sup>, Mark N Brook<sup>1</sup>, Nick Orr<sup>2</sup>, Sunh Kyong Rhie<sup>7</sup>, Elio Riboli<sup>8</sup>, Heather S Feigelson<sup>9</sup>, Loic Le Marchand<sup>10</sup>, Julie E Buring<sup>11</sup>, Diana Eccles<sup>12</sup>, Penelope Miron<sup>13</sup>, Peter A Fasching<sup>14,15</sup>, Hiltrud Brauch<sup>16,17</sup>, Jenny Chang-Claude<sup>18</sup>, Jane Carpenter<sup>19</sup>, Andrew K Godwin<sup>20</sup>, Heli Nevanlinna<sup>21</sup>, Graham G Giles<sup>22,23</sup>, Angela Cox<sup>24</sup>, John L Hopper<sup>25</sup>, Manjeet K Bolla<sup>5</sup>, Qin Wang<sup>5</sup>, Joe Dennis<sup>5</sup>, Ed Dicks<sup>5</sup>, Will J Howat<sup>26</sup>, Nils Schoof<sup>27</sup>, Stig E Bojesen<sup>28</sup>, Diether Lambrechts<sup>29,30</sup>, Annetjie Broeks<sup>6</sup>, Irene L Andrulis<sup>31,32</sup>, Pascal Guénel<sup>33,34</sup>, Barbara Burwinkel<sup>35,36</sup>, Elinor J Sawyer<sup>37</sup>, Antoinette Hollestelle<sup>38</sup>, Olivia Fletcher<sup>2</sup>, Robert Winqvist<sup>39</sup>, Hermann Brenner<sup>40</sup>, Arto Mannermaa<sup>41–43</sup>, Ute Hamann<sup>44</sup>, Alfons Meindl<sup>45,46</sup>, Annika Lindblom<sup>47</sup>, Wei Zheng<sup>48</sup>, Peter Devilee<sup>49,50</sup>, Mark S Goldberg<sup>51,52</sup>, Jan Lubinski<sup>53</sup>, Vessela Kristensen<sup>54,55</sup>, Anthony Swerdlow<sup>1</sup>, Hoda Anton-Culver<sup>56</sup>, Thilo Dörk<sup>57</sup>, Kenneth Muir<sup>58,59</sup>, Keitaro Matsuo<sup>60</sup>, Anna H Wu<sup>7</sup>, Paolo Radice<sup>61,62</sup>, Soo Hwang Teo<sup>63,64</sup>, Xiao-Ou Shu<sup>48</sup>, William Blot<sup>48,65</sup>, Daehee Kang<sup>66</sup>, Mikael Hartman<sup>67,68</sup>, Suleeporn Sangrajrang<sup>69</sup>, Chen-Yang Shen<sup>70,71</sup>, Melissa C Southey<sup>72</sup>, Daniel J Park<sup>72</sup>, Fleur Hammet<sup>72</sup>, Jennifer Stone<sup>25</sup>, Laura J Van't Veer<sup>6</sup>, Emiel J Rutgers<sup>6</sup>, Artitaya Lophatananon<sup>58</sup>, Sarah Stewart-Brown<sup>58</sup>, Pornthep Siriwanarangsana<sup>73</sup>, Julian Peto<sup>74</sup>, Michael G Schrauder<sup>14</sup>, Arif B Ekici<sup>75</sup>, Matthias W Beckmann<sup>14</sup>, Isabel dos Santos Silva<sup>74</sup>, Nichola Johnson<sup>2</sup>, Helen Warren<sup>74</sup>, Ian Tomlinson<sup>76,77</sup>, Michael J Kerin<sup>78</sup>, Nicola Miller<sup>78</sup>, Federick Marme<sup>35,79</sup>, Andreas Schneeweiss<sup>35,79</sup>, Christof Sohn<sup>35</sup>, Therese Truong<sup>33,34</sup>, Pierre Laurent-Puig<sup>80</sup>, Pierre Kerbrat<sup>81</sup>, Børge G Nordestgaard<sup>28</sup>, Sune F Nielsen<sup>28</sup>, Henrik Flyger<sup>82</sup>, Roger L Milne<sup>83</sup>, Jose Ignacio Arias Perez<sup>84</sup>, Primitiva Menéndez<sup>85</sup>, Heiko Müller<sup>40</sup>, Volker Arndt<sup>40</sup>, Christa Stegmaier<sup>86</sup>, Peter Lichtner<sup>87,88</sup>, Magdalena Lochmann<sup>46</sup>, Christina Justenhoven<sup>16,17</sup>, Yon-Dschun Ko<sup>89</sup>, The Gene ENvironmental Interaction and breast Cancer (GENICA) Network<sup>90</sup>, Taru A Muranen<sup>21</sup>, Kristiina Aittomäki<sup>91</sup>, Carl Blomqvist<sup>92</sup>, Dario Greco<sup>21</sup>, Tuomas Heikkinen<sup>21</sup>, Hidemi Ito<sup>60</sup>, Hiroji Iwata<sup>93</sup>, Yasushi Yatabe<sup>94</sup>, Natalia N Antonenkova<sup>95</sup>, Sara Margolin<sup>96</sup>, Vesa Kataja<sup>42,43,97</sup>, Veli-Matti Kosma<sup>41–43</sup>, Jaana M Hartikainen<sup>41–43</sup>, Rosemary Balleine<sup>98,99</sup>, kConFab Investigators<sup>90</sup>, Chiu-Chen Tseng<sup>7</sup>, David Van Den Berg<sup>7</sup>, Daniel O Stram<sup>7</sup>, Patrick Neven<sup>100</sup>, Anne-Sophie Dieudonné<sup>100</sup>, Karin Leunen<sup>100</sup>, Anja Rudolph<sup>18</sup>, Stefan Nickels<sup>18</sup>, Dieter Flesch-Janys<sup>101,102</sup>, Paolo Peterlongo<sup>61,62</sup>, Bernard Peissel<sup>103</sup>, Loris Bernard<sup>104,105</sup>, Janet E Olson<sup>3</sup>, Xianshu Wang<sup>3,106</sup>, Kristen Stevens<sup>3</sup>, Gianluca Severi<sup>22,25</sup>, Laura Baglietto<sup>22,25</sup>, Catriona McLean<sup>107</sup>, Gerhard A Coetzee<sup>7,108</sup>, Ye Feng<sup>7</sup>, Brian E Henderson<sup>7</sup>, Fredrick Schumacher<sup>7</sup>, Natalia V Bogdanova<sup>57,109</sup>, France Labrèche<sup>110</sup>, Martine Dumont<sup>111</sup>, Cheng Har Yip<sup>64</sup>, Nur Aishah Mohd Taib<sup>64</sup>, Ching-Yu Cheng<sup>67,68,112</sup>, Martha Shrubsole<sup>48</sup>, Jirong Long<sup>48</sup>, Katri Pylkäs<sup>39</sup>, Arja Jukkola-Vuorinen<sup>113</sup>, Saila Kauppila<sup>114</sup>,

Julia A Knight<sup>32,115,116</sup>, Gord Glendon<sup>32</sup>, Anna Marie Mulligan<sup>117,118</sup>, Robertus A E M Tollenaar<sup>119</sup>, Caroline M Seynaeve<sup>38</sup>, Mieke Kriege<sup>38</sup>, Maartje J Hooning<sup>38</sup>, Ans M W van den Ouweland<sup>120</sup>, Carolien H M van Deurzen<sup>50</sup>, Wei Lu<sup>121</sup>, Yu-Tang Gao<sup>122</sup>, Hui Cai<sup>48</sup>, Sabapathy P Balasubramanian<sup>24</sup>, Simon S Cross<sup>123</sup>, Malcolm W R Reed<sup>24</sup>, Lisa Signorello<sup>48</sup>, Qiuyin Cai<sup>52</sup>, Mitul Shah<sup>124</sup>, Hui Miao<sup>68</sup>, Ching Wan Chan<sup>125</sup>, Kee Seng Chia<sup>68</sup>, Anna Jakubowska<sup>53</sup>, Katarzyna Jaworska<sup>53</sup>, Katarzyna Durda<sup>53</sup>, Chia-Ni Hsiung<sup>71</sup>, Pei-Ei Wu<sup>126</sup>, Jyh-Cherng Yu<sup>127</sup>, Alan Ashworth<sup>2</sup>, Michael Jones<sup>1</sup>, Daniel C Tessier<sup>128</sup>, Anna González-Neira<sup>129</sup>, Guillermo Pita<sup>129</sup>, M Rosario Alonso<sup>129</sup>, Daniel Vincent<sup>128</sup>, Francois Bacot<sup>128</sup>, Christine B Ambrosone<sup>130</sup>, Elisa V Bandera<sup>131</sup>, Esther M John<sup>132,133</sup>, Gary K Chen<sup>7</sup>, Jennifer J Hu<sup>134,135</sup>, Jorge L Rodriguez-Gil<sup>134,135</sup>, Leslie Bernstein<sup>136</sup>, Michael F Press<sup>137</sup>, Regina G Ziegler<sup>138</sup>, Robert M Millikan<sup>139,179</sup>, Sandra L Deming-Halverson<sup>48</sup>, Sarah Nyante<sup>139</sup>, Sue A Ingles<sup>7</sup>, Quinten Waisfisz<sup>140</sup>, Helen Tsimiklis<sup>141</sup>, Enes Makalic<sup>23,25</sup>, Daniel Schmidt<sup>23,25</sup>, Minh Bui<sup>23,25</sup>, Lorna Gibson<sup>74</sup>, Bertram Müller-Myhsok<sup>142</sup>, Rita K Schmutzler<sup>143,144</sup>, Rebecca Hein<sup>18,145</sup>, Norbert Dahmen<sup>146</sup>, Lars Beckmann<sup>147</sup>, Kirsimari Aaltonen<sup>21,91,92</sup>, Kamila Czene<sup>27</sup>, Astrid Irwanto<sup>148</sup>, Jianjun Liu<sup>148</sup>, Clare Turnbull<sup>1</sup>, Familial Breast Cancer Study (FBCS)<sup>90</sup>, Nazneen Rahman<sup>1</sup>, Hanne Meijers-Heijboer<sup>140</sup>, Andre G Uitterlinden<sup>149</sup>, Fernando Rivadeneira<sup>149</sup>, Australian Breast Cancer Tissue Bank (ABCTB) Investigators<sup>90</sup>, Curtis Olswold<sup>3</sup>, Susan Slager<sup>3</sup>, Robert Pilarski<sup>150</sup>, Foluso Ademuyiwa<sup>151</sup>, Irene Konstantopoulou<sup>152</sup>, Nicholas G Martin<sup>153</sup>, Grant W Montgomery<sup>153</sup>, Dennis J Slamon<sup>15,154</sup>, Claudia Rauh<sup>14</sup>, Michael P Lux<sup>14</sup>, Sebastian M Jud<sup>14</sup>, Thomas Bruning<sup>155</sup>, JoEllen Weaver<sup>156</sup>, Priyanka Sharma<sup>157</sup>, Harsh Pathak<sup>20</sup>, Will Tapper<sup>12</sup>, Sue Gerty<sup>12</sup>, Lorraine Durcan<sup>12</sup>, Dimitrios Trichopoulos<sup>4,158,159</sup>, Rosario Tumino<sup>160</sup>, Petra H Peeters<sup>161</sup>, Rudolf Kaaks<sup>18</sup>, Daniele Campa<sup>18</sup>, Federico Canzian<sup>18</sup>, Elisabete Weiderpass<sup>27,162–164</sup>, Mattias Johansson<sup>165</sup>, Kay-Tee Khaw<sup>166</sup>, Ruth Travis<sup>167</sup>, Françoise Clavel-Chapelon<sup>33,34</sup>, Laurence N Kolonel<sup>110</sup>, Constance Chen<sup>4</sup>, Andy Beck<sup>168,169</sup>, Susan E Hankinson<sup>170,171</sup>, Christine D Berg<sup>172</sup>, Robert N Hoover<sup>138</sup>, Jolanta Lissowska<sup>173</sup>, Jonine D Figueroa<sup>138</sup>, Daniel I Chasman<sup>11</sup>, Mia M Gaudet<sup>174</sup>, W Ryan Diver<sup>174</sup>, Walter C Willett<sup>175</sup>, David J Hunter<sup>4</sup>, Jacques Simard<sup>111</sup>, Javier Benitez<sup>129,176,177</sup>, Alison M Dunning<sup>124</sup>, Mark E Sherman<sup>138</sup>, Georgia Chenevix-Trench<sup>178</sup>, Stephen J Chanock<sup>138</sup>, Per Hall<sup>27</sup>, Paul D P Pharoah<sup>124,181</sup>, Celine Vachon<sup>3,181</sup>, Douglas F Easton<sup>5,181</sup>, Christopher A Haiman<sup>7,181</sup> & Peter Kraft<sup>4,181</sup>

<sup>1</sup>Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, UK. <sup>2</sup>Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK. <sup>3</sup>Mayo Clinic College of Medicine, Mayo Clinic, Rochester, Minnesota, USA. <sup>4</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. <sup>5</sup>Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>6</sup>Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. <sup>7</sup>Department of Preventive Medicine, Keck School of Medicine, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, USA. <sup>8</sup>School of Public Health, Imperial College, London, UK. <sup>9</sup>Kaiser Permanente, Institute for Health Research, Denver, Colorado, USA. <sup>10</sup>Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, Hawaii, USA. <sup>11</sup>Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. <sup>12</sup>Faculty of Medicine, University of Southampton, Southampton, UK. <sup>13</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. <sup>14</sup>Department of Gynecology and Obstetrics, University Breast Center Franconia, University Hospital Erlangen, Erlangen, Germany. <sup>15</sup>Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. <sup>16</sup>Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany. <sup>17</sup>University of Tübingen, Tübingen, Germany. <sup>18</sup>Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>19</sup>Australian Breast Cancer Tissue Bank, University of Sydney at the Westmead Millennium Institute, Westmead, New South Wales, Australia. <sup>20</sup>Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA. <sup>21</sup>Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. <sup>22</sup>Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Victoria, Australia. <sup>23</sup>School of Population Health, The University of Melbourne, Melbourne, Victoria, Australia. <sup>24</sup>Cancer Research UK/Yorkshire Cancer Research Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield, UK. <sup>25</sup>Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Melbourne, Victoria, Australia. <sup>26</sup>Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK. <sup>27</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. <sup>28</sup>Copenhagen General Population Study, Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark. <sup>29</sup>Vesalius Research Center (VRC), VIB, Leuven, Belgium. <sup>30</sup>Department of Oncology, University of Leuven, Leuven, Belgium. <sup>31</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. <sup>32</sup>Ontario Cancer Genetics Network, Fred A. Litwin Center for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. <sup>33</sup>University Paris-Sud, Unité Mixte de Recherche Scientifique (UMRS) 1018, Villejuif, France. <sup>34</sup>INSERM (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), Environmental Epidemiology of Cancer, Villejuif, France. <sup>35</sup>Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany. <sup>36</sup>Molecular Epidemiology Group, DKFZ, Heidelberg, Germany. <sup>37</sup>Division of Cancer Studies, National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Guy's & St. Thomas' National Health Service (NHS) Foundation Trust in partnership with King's College London, London, UK. <sup>38</sup>Department of Medical Oncology, Erasmus University Medical Center—Daniel Den Hoed Cancer Center, Rotterdam, The Netherlands. <sup>39</sup>Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Genetics, Biocenter Oulu, University of Oulu, Oulu University Hospital, Oulu, Finland. <sup>40</sup>Division of Clinical Epidemiology and Aging Research, DKFZ, Heidelberg, Germany. <sup>41</sup>Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland. <sup>42</sup>School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, Kuopio, Finland. <sup>43</sup>Biocenter Kuopio, Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio, Finland. <sup>44</sup>Molecular Genetics of Breast Cancer, DKFZ, Heidelberg, Germany. <sup>45</sup>Division for Gynaecological Tumor Genetics, Clinic of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany. <sup>46</sup>Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany. <sup>47</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. <sup>48</sup>Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Division of Epidemiology, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. <sup>49</sup>Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. <sup>50</sup>Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands. <sup>51</sup>Department of Medicine, McGill University, Montreal, Quebec, Canada. <sup>52</sup>Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, Quebec, Canada. <sup>53</sup>Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. <sup>54</sup>Department of Genetics, Institute for Cancer Research, Oslo University

Hospital, Radiumhospitalet, Oslo, Norway. <sup>55</sup>Faculty of Medicine (Faculty Division Ahus), Universitetet i Oslo, Oslo, Norway. <sup>56</sup>Department of Epidemiology, University of California-Irvine, Irvine, California, USA. <sup>57</sup>Department of Obstetrics and Gynaecology, Hannover Medical School, Hannover, Germany. <sup>58</sup>Warwick Medical School, Warwick University, Coventry, UK. <sup>59</sup>Institute of Population Health, University of Manchester, Manchester, UK. <sup>60</sup>Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. <sup>61</sup>Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale Tumori (INT), Milan, Italy. <sup>62</sup>IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy. <sup>63</sup>Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, University Malaya Cancer Research Institute, University Malaya, Kuala Lumpur, Malaysia. <sup>64</sup>Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya, Kuala Lumpur, Malaysia. <sup>65</sup>International Epidemiology Institute, Rockville, Maryland, USA. <sup>66</sup>Seoul National University College of Medicine, Seoul, Korea. <sup>67</sup>Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. <sup>68</sup>Saw Swee Hock School of Public Health, National University of Singapore, Singapore. <sup>69</sup>National Cancer Institute, Bangkok, Thailand. <sup>70</sup>Colleague of Public Health, China Medical University, Taichong, Taiwan. <sup>71</sup>Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. <sup>72</sup>Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. <sup>73</sup>Ministry of Public Health, Bangkok, Thailand. <sup>74</sup>Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London, UK. <sup>75</sup>Institute of Human Genetics, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany. <sup>76</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>77</sup>Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. <sup>78</sup>Department of Surgery, Clinical Science Institute, University Hospital and National University of Ireland, Galway, Ireland. <sup>79</sup>National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany. <sup>80</sup>INSERM, Université Paris Sorbonne Cité, UMRS 775, Paris, France. <sup>81</sup>Centre Eugène Marquis, Department of Medical Oncology, Rennes, France. <sup>82</sup>Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark. <sup>83</sup>Genetic & Molecular Epidemiology Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. <sup>84</sup>Servicio de Cirugía General y Especialidades, Hospital Monte Naranco, Oviedo, Spain. <sup>85</sup>Servicio de Anatomía Patológica, Hospital Monte Naranco, Oviedo, Spain. <sup>86</sup>Saarland Cancer Registry, Saarbrücken, Germany. <sup>87</sup>Institute of Human Genetics, Technische Universität München, Munich, Germany. <sup>88</sup>Institute of Human Genetics, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany. <sup>89</sup>Department of Internal Medicine, Evangelische Kliniken Bonn, Johanniter Krankenhaus, Bonn, Germany. <sup>90</sup>A list of members is provided in the **Supplementary Note**. <sup>91</sup>Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland. <sup>92</sup>Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland. <sup>93</sup>Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya, Japan. <sup>94</sup>Department of Pathology and Molecular Diagnostics, Aichi Cancer Center Hospital, Nagoya, Japan. <sup>95</sup>N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus. <sup>96</sup>Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. <sup>97</sup>Cancer Center, Kuopio University Hospital, Kuopio, Finland. <sup>98</sup>Western Sydney Local Health District, Westmead Millennium Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia. <sup>99</sup>Nepean Blue Mountains Local Health District, Westmead Millennium Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia. <sup>100</sup>Multidisciplinary Breast Center, University Hospital Gasthuisberg, Department of Oncology, University of Leuven, Leuven, Belgium. <sup>101</sup>Department of Cancer Epidemiology/Clinical Cancer Registry, University Clinic Hamburg-Eppendorf, Hamburg, Germany. <sup>102</sup>Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany. <sup>103</sup>Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS INT, Milan, Italy. <sup>104</sup>Department of Experimental Oncology, Istituto Europeo di Oncologia, Milan, Italy. <sup>105</sup>Cogentech Cancer Genetic Test Laboratory, Milan, Italy. <sup>106</sup>Department of Laboratory Medicine and Pathology, Division of Experimental Pathology, Mayo Clinic, Rochester, Minnesota, USA. <sup>107</sup>Department of Anatomical Pathology, The Alfred Hospital, Melbourne, Victoria, Australia. <sup>108</sup>Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. <sup>109</sup>Department of Radiation Oncology, Hannover Medical School, Hannover, Germany. <sup>110</sup>Département de Médecine Sociale et Préventive, Département de Santé Environnementale et Santé au Travail, Université de Montréal, Montreal, Quebec, Canada. <sup>111</sup>Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Québec and Laval University, Québec City, Quebec, Canada. <sup>112</sup>Singapore Eye Research Institute, National University of Singapore, Singapore. <sup>113</sup>Department of Oncology, Oulu University Hospital, University of Oulu, Oulu, Finland. <sup>114</sup>Department of Pathology, Oulu University Hospital, University of Oulu, Oulu, Finland. <sup>115</sup>Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. <sup>116</sup>Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. <sup>117</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. <sup>118</sup>Department of Laboratory Medicine, Keenan Research Centre of the Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada. <sup>119</sup>Department of Surgical Oncology, Leiden University Medical Center, Leiden, The Netherlands. <sup>120</sup>Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. <sup>121</sup>Shanghai Center for Disease Control and Prevention, Shanghai, China. <sup>122</sup>Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. <sup>123</sup>Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK. <sup>124</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. <sup>125</sup>Department of Surgery, National University Health System, Singapore. <sup>126</sup>Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. <sup>127</sup>Department of Surgery, Tri-Service General Hospital, Taipei, Taiwan. <sup>128</sup>McGill University and Génomique Québec Innovation Centre, Montreal, Québec, Canada. <sup>129</sup>Human Genotyping Unit—CEGEN, Human Cancer Genetics Programme, CNIO, Madrid, Spain. <sup>130</sup>Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York, USA. <sup>131</sup>The Cancer Institute of New Jersey, New Brunswick, New Jersey, USA. <sup>132</sup>Cancer Prevention Institute of California, Fremont, California, USA. <sup>133</sup>Department of Health Research and Policy, Division of Epidemiology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA. <sup>134</sup>Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, Florida, USA. <sup>135</sup>Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, Florida, USA. <sup>136</sup>Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, Duarte, California, USA. <sup>137</sup>Department of Pathology, Keck School of Medicine, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, USA. <sup>138</sup>Epidemiology and Biostatistics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA. <sup>139</sup>Department of Epidemiology, Gillings School of Global Public Health, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, USA. <sup>140</sup>Section of Oncogenetics, Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands. <sup>141</sup>Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. <sup>142</sup>Statistical Genetics Research Group, Max Planck Institute of Psychiatry, Munich, Germany. <sup>143</sup>Centre of Hereditary Breast and Ovarian Cancer, University Hospital, Cologne, Germany. <sup>144</sup>Centre of Integrated Oncology, University Hospital, Cologne, Germany. <sup>145</sup>PMV (Primärmedizinische Versorgung) Research Group, Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany. <sup>146</sup>Department of Psychiatry, University of Mainz, Mainz, Germany. <sup>147</sup>Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany. <sup>148</sup>Human Genetics Division, Genome Institute of Singapore, Singapore. <sup>149</sup>Department of Internal Medicine and Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>150</sup>Department of Internal Medicine, James Comprehensive Cancer Center, Ohio State University, Columbus, Ohio, USA. <sup>151</sup>Roswell Park Cancer Institute, Buffalo, New York, USA. <sup>152</sup>Molecular Diagnostics Laboratory, Institute of Radioisotopes and Radiodiagnostic Products (IRRP), National Centre for Scientific Research Demokritos, Aghia Paraskevi Attikis, Athens, Greece. <sup>153</sup>QIMR GWAS Collective, Queensland Institute of Medical Research, Brisbane, Queensland, Australia. <sup>154</sup>Department of Medicine, Division of Hematology and Oncology, University of California, Los Angeles, Los Angeles, California, USA. <sup>155</sup>Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, Germany. <sup>156</sup>Biosample Repository, Fox Chase Cancer Center, Philadelphia, Pennsylvania, USA. <sup>157</sup>Division of Hematology and Oncology, Department of Internal Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA. <sup>158</sup>Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece. <sup>159</sup>Hellenic Health Foundation, Athens, Greece. <sup>160</sup>Cancer Registry, Histopathology Unit Civile MPArezzo Hospital, Ragusa, Italy. <sup>161</sup>Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands. <sup>162</sup>Department of Community Medicine, University of Tromsø, Tromsø, Norway. <sup>163</sup>Folkhälsan Research Cancer Centre, Helsinki, Finland. <sup>164</sup>Cancer Registry of Norway, Oslo, Norway. <sup>165</sup>Genetic Epidemiology Group, International Agency for Research on Cancer, World Health Organization, Lyon, France. <sup>166</sup>Clinical Gerontology Unit, University of Cambridge, Cambridge, UK. <sup>167</sup>Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK. <sup>168</sup>Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA. <sup>169</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. <sup>170</sup>Division of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, Massachusetts, USA. <sup>171</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. <sup>172</sup>Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland, USA. <sup>173</sup>Department of Cancer Epidemiology and Prevention, M Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland. <sup>174</sup>Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA. <sup>175</sup>Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. <sup>176</sup>Human Genetics Group, CNIO, Madrid, Spain. <sup>177</sup>Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain. <sup>178</sup>Department of Genetics, Queensland Institute of Medical Research, Brisbane, Queensland, Australia. <sup>179</sup>Deceased. <sup>180</sup>These authors contributed equally to this work. <sup>181</sup>These authors jointly directed this work. Correspondence should be addressed to M.G.-C. ([montse.garciaclosas@icr.ac.uk](mailto:montse.garciaclosas@icr.ac.uk)).



## ONLINE METHODS

**ER-negative breast cancer GWAS.** Three GWAS of ER-negative breast cancer were conducted in populations of European ancestry by National Cancer Institute (NCI) BPC3 (refs. 7,18), TNBCC<sup>5,6</sup> and C-BCAC.

ER-negative status for BPC3 and C-BCAC cases was determined from review of medical records or state cancer registry information. TNBCC focused on triple-negative cases, defined as individuals with ER-negative, PR-negative and HER2-negative breast cancer using data from medical records<sup>5,6</sup>. The BPC3 GWAS included 2,188 ER-negative cases and 26,477 controls from 8 studies (CPSII, EPIC, MEC, NHS, NHSII, PLCO, PBCS and WGHS), genotyped using different versions of Illumina SNP arrays<sup>7,18</sup>. A total of 1,718 triple-negative cases from 11 studies (ABCTB, BBCC, DFCL, FCCC, GENICA, HEBCS, MARIE, MCBCS, MCCS, POSH and SBCS) were genotyped for the TNBCC GWAS using Illumina SNP arrays<sup>5</sup>. Data for TNBCC controls ( $N = 3,670$ ) were obtained from a Finnish study (HEBCS) and publicly available controls of European ancestry from the United States (CGEMS), Germany (KORA), Australia (QIMR) and the UK (Wellcome Trust Case Control Consortium 2, WTCCC2) genotyped using Illumina arrays<sup>5</sup>. Samples from the four latter studies are not counted in the total number of TNBCC studies because they only provided controls for other studies. C-BCAC performed a meta-analysis of 9 GWAS that included data on 10,052 breast cancer cases and 12,575 controls<sup>8</sup>. Five studies (ABCFS, MARIE, HEBCS, SASBAC and UK2) provided data on ER status from medical records or cancer registries and contributed data on 702 ER-negative cases and 7,713 controls of European ancestry. All C-BCAC studies were genotyped with versions of Illumina arrays. Control data for C-BCAC were obtained from individual studies or publicly available data.

Standard genotyping quality control procedures were performed for each GWAS as previously described<sup>5,7,8</sup>. Estimated per-allele log(OR) and standard error were calculated for each SNP using unconditional logistic regression on allele counts (dosages), as implemented in ProbABEL<sup>19</sup>. Analyses were adjusted by study, country of origin or principal components as previously described<sup>5,7,8</sup>. Analyses assumed a log-additive genetic model, and  $P$  values were based on the 1-degree-of-freedom Wald test. Quantile-quantile plots from each GWAS showed no substantial evidence for cryptic population substructure or differential genotype calling between cases and controls. The estimated inflation factor ( $\lambda$ ) was 1.02 for BPC3 (ref. 7), 1.04 for TNBCC<sup>6</sup> and 0.98 for C-BCAC (Supplementary Fig. 1).

SNPs were selected for the iCOGS custom genotyping array separately by each participating group (see details in Michailidou *et al.*<sup>8</sup>). BPC3 nominated independent SNPs with a 1-degree-of-freedom log-additive trend test  $P < 0.02$  or with  $P < 0.02$  for one of several auxiliary tests, including tests for dominant or recessive effects of the minor allele and case-only tests comparing PR-positive to PR-negative tumors. SNPs from C-BCAC were selected on the basis of the 1-degree-of-freedom trend test for ER-negative disease. TNBCC nominated SNPs on the basis of log-additive trend test  $P < 0.01$ . Subsequent analyses that combined OR estimates across GWAS and follow-up samples only included SNPs that had been directly genotyped on the iCOGS array and had passed genotyping quality control. SNPs successfully genotyped on iCOGS but not included on the chips used for the GWAS were imputed within each GWAS before combining results with iCOGS data. Imputation was performed within each study and genotyping array using the HapMap Phase 2 CEU reference panel and MACH software package v1.0. SNPs with low imputation quality ( $r^2 < 0.3$ ) or minor allele frequency (MAF)  $< 1\%$  were excluded.

**iCOGS genotyping.** Samples for follow-up analyses were drawn from 50 studies participating in BCAC (40 from populations of predominantly European ancestry (including CTS, DEMOKRITOS, NBCS, NBHS, OSUCCG, RPCI and SKKDKFZS from TNBCC), 9 of Asian ancestry and 1 of African-American ancestry) with information on ER status. Most breast cancer cases in BCAC studies have not been tested for *BRCA1* mutations; however, the frequency of mutations in the studied populations is expected to be low. Samples were genotyped as part of the COGS Project using a custom Illumina Infinium array (iCOGS) at four genotyping centers (Supplementary Table 1). The most common source of data for ER, PR and HER2 status was medical records, followed by immunohistochemistry performed on tumor tissue microarrays (TMAs) or whole-section tumor slides. Breast cancer cases in the BCAC follow-up with

missing data on ER status and cases from one study (PBCS) that included only ER-positive cases are excluded from this report. Studies were required to provide ~2% of samples in duplicate.

The iCOGS chip included a total of 211,155 SNPs selected in collaboration with other consortia of *BRCA1* and *BRCA2* mutation carriers (CIMBA), ovarian cancer (OCAC) and prostate cancer (PRACTICAL). Genotype calling and quality control analyses were conducted by a single analysis center at the University of Cambridge<sup>8</sup>. A total of 13,276 SNPs proposed by the combined ER-negative GWAS yielded high-quality genotype data (5,738 from BPC3, 4,628 from TNBCC and 2,910 from C-BCAC).

**Statistical analysis.** After quality control exclusions<sup>8</sup>, BCAC follow-up data were analyzed using the Genotype Library and Utilities (GLU) package to estimate per-allele ORs and standard errors for each SNP using unconditional logistic regression. Analyses were stratified by ancestry (European, Asian or African). For samples of European ancestry, BCAC follow-up analyses were adjusted for seven principal components (the first six plus an additional component to reduce inflation for the LMBC study).

GWAS and BCAC follow-up results were combined using inverse variance-weighted fixed-effects meta-analysis, as implemented in METAL<sup>20</sup>. Forest plots showing study-specific estimates and fixed-effects meta-analysis for SNPs showing genome-wide significance were drawn using the command `metan` in STATA v.12. Samples that overlapped among the three GWAS and the BCAC follow-up were identified by concordance of genotypes and removed from either the GWAS or follow-up data set before this analysis so that each data set contributing to the meta-analysis was independent of the others (see Supplementary Table 1 for the counts of case and control included in the analyses after removing overlapping samples). Heterogeneity by study was evaluated using the  $Q$  statistic.

Analyses in this report focused first on the 13,276 SNPs proposed by the ER-negative breast cancer GWAS. For SNPs showing evidence of association with ER-negative breast cancer at  $P < 1 \times 10^{-6}$ , we also evaluated correlated SNPs in the rest of COGS and reported on the most significant SNP in the region. For the regions that reached genome-wide statistical significance ( $P < 5 \times 10^{-8}$ ), we performed additional analyses examining heterogeneity in the associated effect by tumor type and subject characteristics using the most significant SNP in the region. The associations between these SNPs and ER-positive breast cancer were assessed using 25,227 ER-positive cases of European ancestry in BCAC who had been genotyped as part of the COGS Project. Differences in the strength of the associations with ER-positive and ER-negative breast cancers were assessed using case-only analyses (Supplementary Table 5). Stratum-specific estimates of per-allele OR by categories of age and family history of disease were obtained from logistic regression models (Supplementary Tables 6 and 7), and differences in ORs across strata were tested using an ordinal-product interaction term.

We also assessed associations between the most significant markers and ER-negative breast cancer in Asian and African-American populations. The Asian-ancestry analyses included 1,547 ER-negative cases and 6,624 controls in 9 studies from BCAC. The African-American analyses included 91 ER-negative cases and 252 controls in 1 BCAC study and 988 ER-negative cases and 2,745 controls in 9 studies from AABC<sup>5</sup> (Supplementary Table 1). Both the Asian-ancestry and African-American analyses adjusted for the first two principal components of genetic variation, calculated separately in each ancestry group. Differences by ancestry were tested by a  $\chi^2$  test comparing summary ORs across the three ancestry groups.

**Bioinformatics.** In an attempt to identify functionality in regions of interest, we used the open-source R/Bioconductor package FunciSNP version 0.1.14 (Functional Integration of SNPs)<sup>21</sup> (S.K.R., S.G. Coetzee, H. Noushmehr, C. Yan, J.M. Kim *et al.*, unpublished data), which systematically integrates 1000 Genomes Project SNP data (June 2011 data release) with chromatin features of interest. For each of the four newly associated ER-negative breast cancer markers we analyzed all SNPs within a 1-Mb window that were in LD ( $r^2 > 0.5$ ) with the index marker (according to the 1000 Genome Project CEU panel). We assessed whether these SNPs colocalized with 13 different chromatin features that capture open chromatin regions and enhancers across the genome, using data generated by next-generation



sequencing technologies. Information on open chromatin states (H3K9ac and H3K14ac), nucleosome-depleted regions (DNase I and FAIRE), enhancers (H3K4me1) and active/engaged enhancers (H3K27ac) was either generated by the Coetzee Laboratory (S.K.R. *et al.*, unpublished data) or harvested from the Encyclopedia of DNA Elements (ENCODE) Project. All chromatin features were identified in normal human mammary epithelial cells (HMECs) and triple-negative breast cancer cells (MDA-MB-231). We used the UCSC Genome Browser (see URLs) with potentially functional SNPs identified using FunciSNP and chromatin features tracks to generate images (Supplementary Fig. 5).

**Ethics.** All women in participating studies provided written consent for the research, and approval for the study was obtained from the local ethical

review board relevant to each institution. Collection of blood samples and clinical data from subjects was performed in accordance with local guidelines and regulations.

18. 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).
19. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
20. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
21. Coetzee, S.G., Rhie, S.K., Berman, B.P., Coetzee, G.A. & Noushmehr, H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic Acids Res.* **40**, e139 (2012).