

Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes

Genome-wide association studies have identified breast cancer risk variants in over 150 genomic regions, but the mechanisms underlying risk remain largely unknown. These regions were explored by combining association analysis with *in silico* genomic feature annotations. We defined 205 independent risk-associated signals with the set of credible causal variants in each one. In parallel, we used a Bayesian approach (PAINTOR) that combines genetic association, linkage disequilibrium and enriched genomic features to determine variants with high posterior probabilities of being causal. Potentially causal variants were significantly over-represented in active gene regulatory regions and transcription factor binding sites. We applied our INQUISIT pipeline for prioritizing genes as targets of those potentially causal variants, using gene expression (expression quantitative trait loci), chromatin interaction and functional annotations. Known cancer drivers, transcription factors and genes in the developmental, apoptosis, immune system and DNA integrity checkpoint gene ontology pathways were over-represented among the highest-confidence target genes.

Genome-wide association studies (GWASs) have identified genetic variants associated with breast cancer risk in more than 150 genomic regions^{1,2}. However, the variants and genes driving these associations are mostly unknown, with fewer than 20 regions studied in detail^{3–20}. Here, we aimed to fine-map all known breast cancer susceptibility regions using dense genotype data on >217,000 subjects participating in the Breast Cancer Association Consortium (BCAC) and the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA). All samples were genotyped using the OncoArray^{1,2,21} or the iCOGS chip^{22,23}. Stepwise multinomial logistic regression was used to identify independent association signals in each region and to define credible causal variants (CCVs) within each signal. We found genomic features significantly overlapping the CCVs. We then used a Bayesian approach, integrating genomic features and genetic associations, to refine the set of likely causal variants and calculate their posterior probabilities. Finally, we integrated genetic and *in silico* epigenetic expression and chromatin conformation data to infer the likely target genes of each signal.

Results

Most breast cancer genomic regions contain multiple independent risk-associated signals. We included 109,900 cases of breast cancer and 88,937 controls, all of European ancestry, from 75 studies in the BCAC. Genotypes (directly observed or imputed) were available for 639,118 single nucleotide polymorphisms (SNPs), deletions/insertions and copy number variants (CNVs) with a minor allele frequency (MAF) $\geq 0.1\%$ within 152 previously defined, risk-associated regions (Supplementary Table 1 and Fig. 1). Multivariate logistic regression confirmed associations for 150 out of 152 regions at a significance threshold of $P < 10^{-4}$ (Supplementary Table 2a). To determine the number of independent risk signals within each region, we applied stepwise multinomial logistic regression, deriving the association of each variant, conditional on the more significant ones, in order of statistical significance. Finally, we defined CCVs in each signal as variants with conditional P values within two orders of magnitude of the index variant²⁴. We classified the evidence for each independent signal, and its CCVs, as either strong (conditional $P < 10^{-6}$) or moderate ($10^{-6} < \text{conditional } P < 10^{-4}$).

From the 150 genomic regions, we identified 352 independent risk signals containing 13,367 CCVs, 7,394 of which were within the 196 strong-evidence signals across 129 regions (Fig. 2a,b). The number of signals per region ranged from 1–11, with 79 (53%) containing multiple signals. We noted a wide range of CCVs per signal, but in 42 signals there was only a single CCV: for these signals, the simplest hypothesis is that the CCV was causal (Fig. 2c,d and Table 1). Furthermore, within signals with few CCVs (< 10), the mean scaled combined annotation-dependent depletion score was higher than in signals with more CCVs (13.1 versus 6.7 for CCVs in exons; $P_{t\text{-test}} = 2.7 \times 10^{-4}$), suggesting that these are more likely to be functional.

The majority of breast tumors express the estrogen receptor (ER positive), but ~20% do not (ER negative); these two tumor types have distinct biological and clinical characteristics²⁵. Using a case-only analysis for the 196 strong-evidence signals, we found 66 signals (34%; containing 1,238 CCVs) where the lead variant conferred a greater relative risk of developing ER-positive tumors (false discovery rate (FDR) = 5%), and 29 (15%; 646 CCVs) where the lead variant conferred a greater risk of ER-negative cancer tumors (FDR = 5%) (Supplementary Table 2b and Fig. 2e). The remaining 101 signals (51%; 5,510 CCVs) showed no difference by ER status (referred to as ER neutral).

Patients with *BRCA1* mutations are more likely to develop ER-negative tumors²⁶. Hence, to increase our power to identify ER-negative signals, we performed a fixed-effects meta-analysis, combining association results from *BRCA1* mutation carriers in CIMBA with the BCAC ER-negative association results. This meta-analysis identified ten additional signals (seven ER-negative and three ER-neutral), making 206 strong-evidence signals (17% ER negative) containing 7,652 CCVs in total (Fig. 2f). More than one-quarter of the CCVs (2,277) were accounted for by one signal, resulting from strong linkage disequilibrium with a CNV. The remaining analyses focused on the other 205 strong signals across 128 regions (Supplementary Table 2c).

The proportion of the familial relative risk (FRR) of breast cancer explained by all 206 strong signals was 20.6%, compared with 17.6% when only the lead SNP for each region was considered. The proportion of the FRR explained increased by a further 3% (to 23.6%) when all 352 signals were considered (Supplementary Table 2d).

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

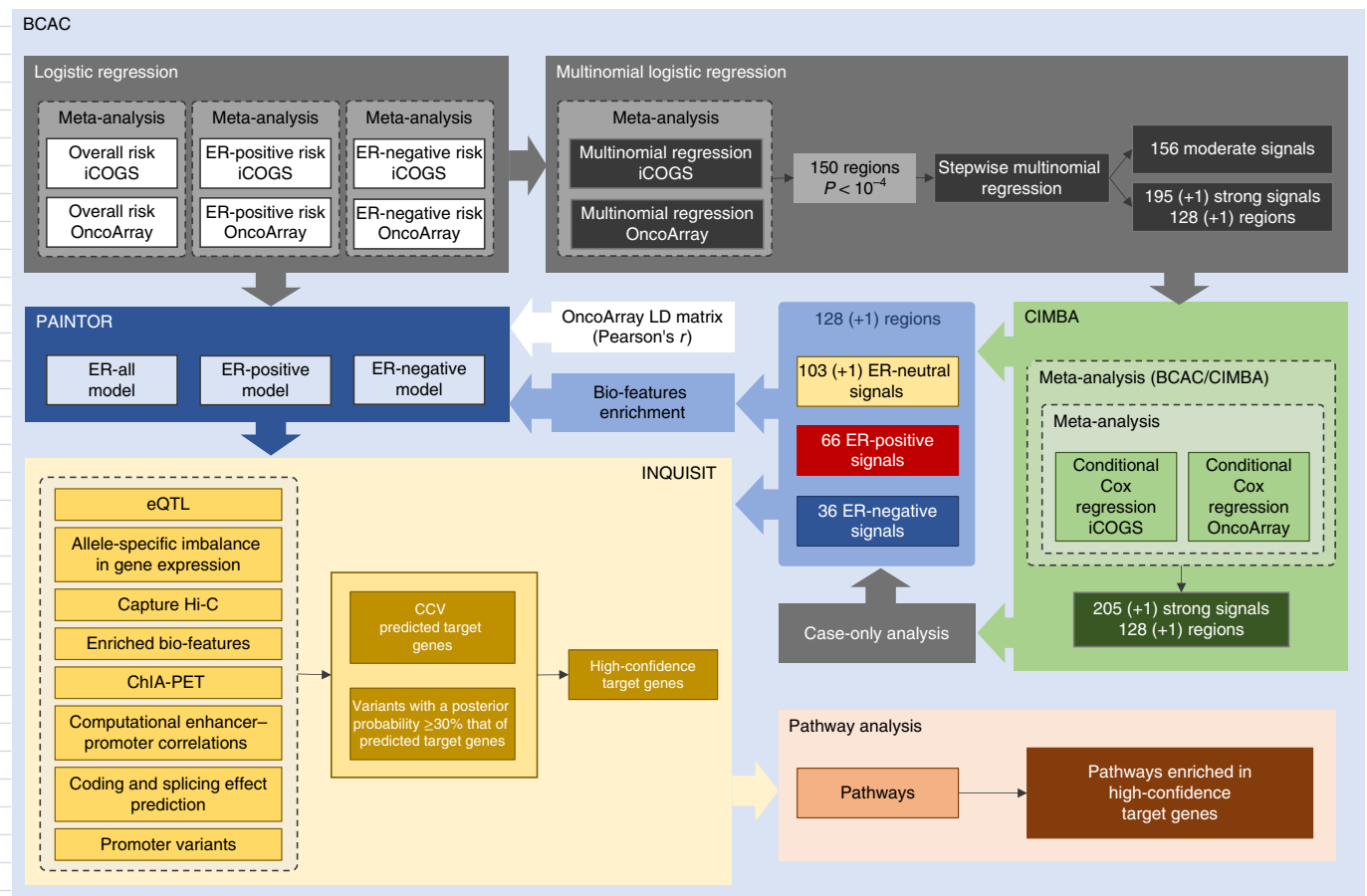


Fig. 1 | Flowchart summarizing the study design. Logistic regression summary statistics were used to select the final set of variants to run stepwise multinomial regression. These results were meta-analyzed with CIMBA to provide the final set of strong independent signals and their CCVs. Through case-only analysis, we identified significant differences in effect sizes between ER-positive and ER-negative breast cancer and used this to classify the phenotype for each independent signal. With these strong CCVs, we ran the bio-features enrichment analysis, which identified the features to be included in the PAINTOR models, together with the OncoArray logistic regression summary statistics and the OncoArray linkage disequilibrium. Both multinomial regression CCVs and PAINTOR high-posterior-probability (PP) variants were analyzed with INQUISIT to determine high-confidence target genes. Finally, we used the set of high-confidence target genes to identify enriched pathways. iCOGS and OncoArray Cox regression was conditional on the index variants from BCAC strong signals.

CCVs are over-represented in active gene regulatory regions and transcription factor binding sites (TFBSs). We constructed a database of mapped genomic features in seven primary cells derived from normal breast and 19 breast cell lines using publicly available data, resulting in 811 annotation tracks in total. These ranged from general features (such as whether a variant was in an exon or in open chromatin) to more specific features (such as cell-specific transcription factor binding or histone marks (determined through chromatin immunoprecipitation followed by sequencing (ChIP-Seq) experiments) in breast-derived cells or cell lines). Using logistic regression, we examined the overlap of these genomic features with the positions of 5,117 CCVs in the 195 strong-evidence BCAC signals versus the positions of 622,903 variants excluded as credible candidates in the same regions (Supplementary Fig. 1a and Supplementary Table 3). We found significant enrichment of CCVs (FDR=5%) in four genomic features (open chromatin, actively transcribed genes, gene regulatory regions and binding sites), as described below.

Open chromatin. As shown in Fig. 3a, DNase I hypersensitive sites sequencing and formaldehyde-assisted isolation of regulatory elements sequencing showed significant enrichment of CCVs in open chromatin in ER-positive breast cancer cell lines and normal breast.

Conversely, we found depletion of CCVs within heterochromatin (determined by the H3K9me3 mark in normal breast, and by chromatin state in ER-positive cells²⁷).

Actively transcribed genes. Significant enrichment of CCVs was also found in actively transcribed genes in normal breast and ER-positive cell lines (as defined by H3K36me3 or H3K79me2 histone marks; Fig. 3a). Enrichment was larger for ER-neutral CCVs than for those affecting either ER-positive or ER-negative tumors.

Gene regulatory regions. CCVs overlapped distal gene regulatory elements in ER-positive breast cancer cell lines (defined by H3K4me1 or H3K27ac marks; Fig. 3b). This was confirmed using the Encyclopedia of DNA Elements (ENCODE) definition of active enhancers in MCF-7 cells (enhancer-like regions defined by combining DNase and H3K27ac marks), as well as the definition of refs.^{27,28} (Supplementary Table 3). Under these more stringent definitions, enrichment among ER-positive CCVs was significantly larger than ER-negative or ER-neutral CCVs. Data from ref.²⁷ showed that 73% of active enhancer regions overlapped by ER-positive CCVs in ER-positive cells (MCF-7) are inactive in the normal human mammary epithelial (HMEC) breast cell line; thus, these enhancers appear to be MCF-7 specific.

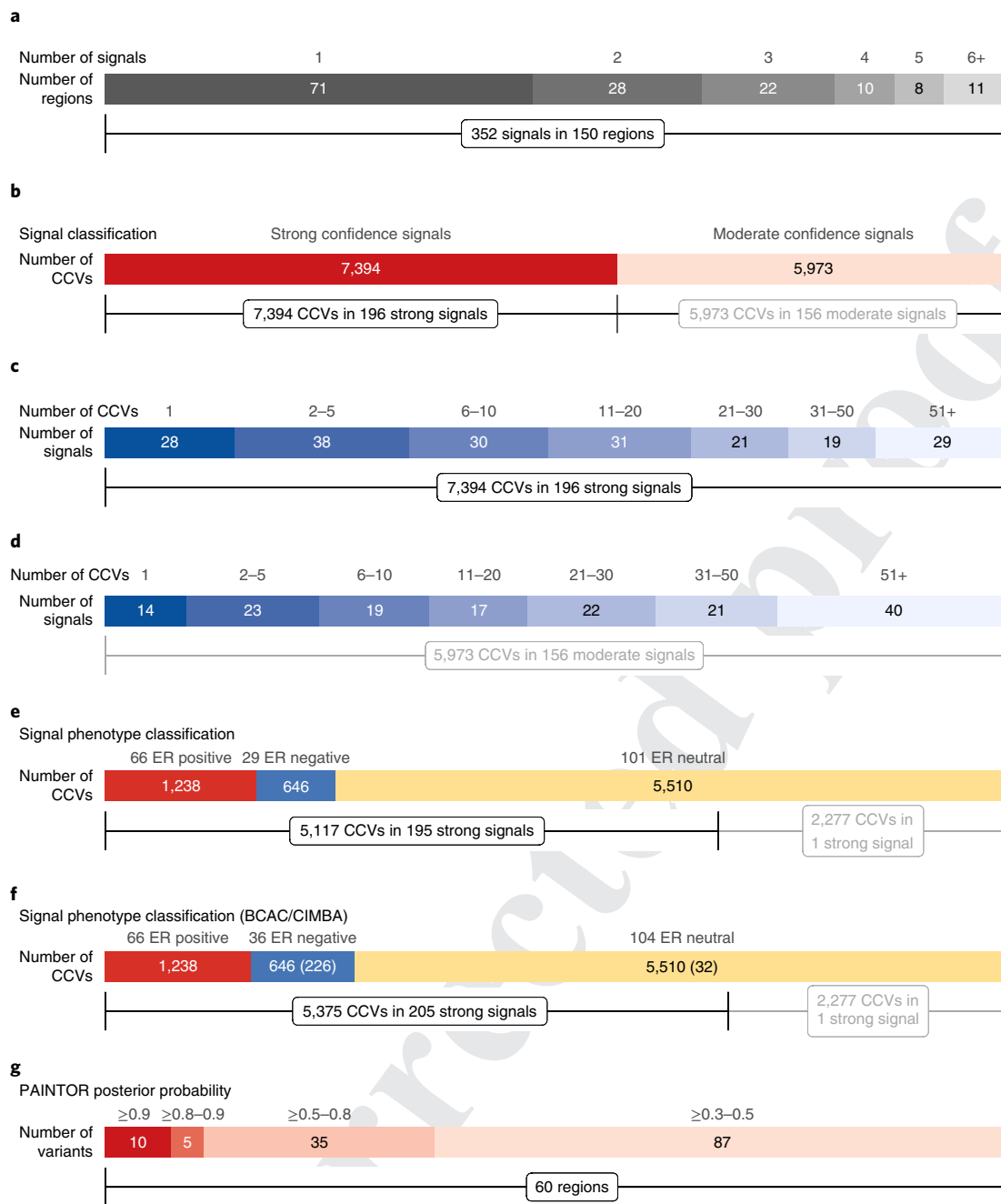


Fig. 2 | Determining independent risk signals and CCVs. a, Number of independent signals per region, identified through multinomial stepwise logistic regression. **b**, Signal classification as strong- or moderate-confidence signals. **c,d**, Number of CCVs per signal in strong- (**c**) and moderate-confidence signals (**d**), identified through multinomial stepwise logistic regression. **e**, Subtype classification of strong signals into ER positive, ER negative and signals equally associated with both phenotypes (ER neutral) from the BCAC analysis. **f**, Subtype classification from the meta-analysis of BCAC and CIMBA. Numbers in brackets show the numbers of CCVs from the meta-analysis of BCAC and CIMBA. **g**, Number of variants at different posterior probability thresholds. In total, 15 variants reached a posterior probability of $\geq 80\%$ by at least one of the three models (ER all, ER positive or ER negative).

We also detected significant enrichment of CCVs in active promoters in ER-positive cells (defined by H3K4me3 marks in T-47D), although the evidence for this effect was weaker than for distal regulatory elements (defined by H3K27ac marks in MCF-7; Fig. 3b). Only ER-positive CCVs were significantly enriched in T-47D active promoters. Conversely, CCVs were depleted among repressed gene regulatory elements (defined by H3K27me3 marks) in normal breast (Fig. 3b). As a control, we performed similar analyses with autoimmune disease CCVs²⁹ (Methods) and relevant B and T cells (Fig. 3b–e). The strongest evidence of enrichment of breast cancer

CCVs was found at regulatory regions active in ER-positive cells (Fig. 3b), whereas enrichment of autoimmune CCVs was in regulatory regions active in B and T cells (Fig. 3e). We also compared the enrichment of our CCVs in enhancer-like and promoter-like regions (defined by ENCODE; Supplementary Fig. 1b). The strongest evidence of enrichment of ER-positive CCVs in enhancer-like regions was found in MCF-7 cells—the only ER-positive cell line in ENCODE (Supplementary Fig. 1b). These results highlight both the tissue specificity and disease specificity of these histone-marked gene regulatory regions.

Binding sites. We observed significant enrichment of CCVs in the binding sites for 40 TFBSs determined by ChIP-Seq (Fig. 3f–h). The majority of the experiments were performed in ER-positive cell lines (90 TFBSs; 20 with data in ER-negative cell lines, 76 with data in ER-positive cell lines and 16 with data in normal breast). These TFBSs overlap each other and histone marks of active regulatory regions (Supplementary Fig. 2). Enrichment in five TFBSs (*ESR1*, *FOXA1*, *GATA3*, *TCF7L2* and *E2F1*) has been reported previously^{2,30}. All 40 TFBSs were significantly enriched in ER-positive CCVs (Fig. 3f), seven were also enriched in ER-negative CCVs and nine were enriched in ER-neutral CCVs (Fig. 3g–h). *ESR1*, *FOXA1*, *GATA3* and *EP300* TFBSs were enriched in all CCV ER subtypes. However, the enrichment for *ESR1*, *FOXA1* and *GATA3* was stronger for ER-positive CCVs than for ER-negative or ER-neutral CCVs.

CCVs significantly overlap consensus transcription factor binding motifs. We investigated whether CCVs were also enriched within consensus transcription factor binding motifs by conducting a motif search within active regulatory regions (ER-positive CCVs at H3K4me1 marks in MCF-7). We identified 30 motifs from eight transcription factor families, with enrichment in ER-positive CCVs (FDR = 10%; Supplementary Table 4a) and a further five motifs depleted among ER-positive CCVs. To assess whether the motifs appeared more frequently than by chance at active regulatory regions overlapped by our ER-positive CCVs, we compared motif presence in a set of randomized control sequences (Methods). Thirteen of 30 motifs were more frequent at active regulatory regions with ER-positive CCV enrichment; these included seven homeodomain motifs and two forkhead factors (Supplementary Table 4b).

When we looked at the change in predicted binding affinity, 57 ER-positive signals (86%) included at least one CCV predicted to modify the binding affinity of the enriched TFBSs (at least twofold; Supplementary Table 4c). Forty-eight ER-positive signals (73%) had at least one CCV predicted to modify the binding affinity greater than tenfold. This analysis validates previous reports of breast cancer causal variants that alter the DNA binding affinity for *FOXA1* (refs. ^{3,30}).

Bayesian fine-mapping incorporating functional annotations and linkage disequilibrium. As an alternative statistical approach for inferring likely causal variants, we applied PAINTOR³¹ to the same 128 regions (Fig. 1). In brief, PAINTOR integrates genetic association results, linkage disequilibrium structure and enriched genomic features in an empirical Bayes framework and derives the posterior probability of each variant being causal, conditional on available data. To eliminate artefacts due to differences in genotyping and imputation across platforms, we restricted PAINTOR analyses to cases and controls typed using OncoArray (61% of the total). We identified seven variants with a high posterior probability (HPP $\geq 80\%$) of being causal for overall breast cancer, and ten for the ER-positive subtype (Table 1); two of these had a HPP $> 80\%$ for both ER-positive and overall breast cancer. These 15 HPP variants (HPPVs; $\geq 80\%$) were distributed across 13 regions. We also identified an additional 35 variants in 25 regions with HPP (≥ 50 and $< 80\%$) for ER-positive, ER-negative or overall breast cancer (Fig. 2g).

Consistent with the CCV analysis, we found evidence that most regions contained multiple HPPVs; the sum of posterior probabilities across all variants in a region (an estimate of the number of distinct causal variants in the region) was > 2.0 for 84 out of 86 regions analyzed for overall breast cancer, with a maximum of 16.1 and a mean of 6.4. For ER-positive cancer, 46 out of 47 regions had total posterior probability of > 2.0 (maximum: 18.3; mean: 6.5). For ER-negative cancer, 17 out of 23 regions had a total posterior probability of > 2.0 (maximum: 9.1; mean: 3.2).

Although for many regions we were not able to identify HPPVs, we were able to reduce the proportion of variants needed to account for 80% of the total posterior probability in a region to $< 5\%$ for 65 regions for overall breast cancer, 43 regions for ER-positive breast cancer and 18 regions for ER-negative breast cancer (Supplementary Fig. 3a–c). PAINTOR analyses were also able to reduce the set of likely causal variants in many cases. After summing the posterior probabilities for CCVs in each of the overall breast cancer signals, 39 out of 100 strong-evidence signals had a total posterior probability of > 1.0 . The number of CCVs in these signals ranged from 1–375 (median: 24), but the number of variants needed to capture 95% of the total posterior probability in each signal ranged from 1–115 (median: 12), representing an average reduction of 43% in the number of variants needed to capture the signal.

PAINTOR and CCV analyses were generally consistent, yet complementary. Only 3.3% of variants outside of the set of strong-signal CCVs for overall breast cancer had a posterior probability of $> 1\%$, and only 48 (0.013%) of these had a posterior probability of $> 30\%$ (Supplementary Fig. 3d). At ER-positive and ER-negative signals, respectively, 3.1 and 1.6% of the non-CCVs at strong signals had a posterior probability of $> 1\%$, and 40 (0.019%) and 3 (0.003%) of these had a posterior probability of $> 30\%$ (Supplementary Fig. 3e–f). For the non-CCVs at strong-evidence signals with a posterior probability of $> 30\%$, the relatively HPP may be driven by the addition of functional annotation. Indeed, the incorporation of functional annotations more than doubled the posterior probability for 64 out of 88 variants when compared with a PAINTOR model with no functional annotations.

CCVs co-localize with variants controlling local gene expression. We used four breast-specific expression quantitative trait loci (eQTL) datasets to identify a credible set of variants associated with differences in gene expression (expression variants): tumor tissue from the Nurses' Health Study (NHS)³² and The Cancer Genome Atlas (TCGA)³³; and normal breast tissue from the NHS and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)³⁴. We then examined the overlap of expression variants (for each gene, expression variants were defined as those variants that had a *P* value within two orders of magnitude of the variant most significantly associated with that gene's expression) with CCVs (Methods). There was significant overlap of CCVs with expression variants from both the NHS normal and breast cancer tissue studies (normal breast: odds ratio (OR) = 2.70; $P = 1.7 \times 10^{-5}$; tumor tissue: OR = 2.34; $P = 2.6 \times 10^{-4}$; Supplementary Table 3). ER-neutral CCVs overlapped with expression variants in normal tissue more frequently than ER-positive and ER-negative CCVs ($OR_{ER\text{-neutral}} = 3.51$; $P = 1.3 \times 10^{-5}$). Cancer risk CCVs overlapped credible expression variants in 128 out of 205 signals (62%) in at least one of the datasets (Supplementary Table 5a,b). Sixteen additional variants with a posterior probability of $\geq 30\%$, not included among the CCVs, also overlapped with a credible expression variant (Supplementary Table 5a,b).

Transcription factors and known somatic breast cancer drivers are over-represented among prioritized target genes. We assumed that causal variants function by affecting the behavior of a local target gene. However, it is challenging to define target genes or to determine how they may be affected by the causal variant. Few potentially causal variants directly affect protein coding: we observed 67 out of 5,375 CCVs and 19 out of 137 HPPVs ($\geq 30\%$) in protein-coding regions. Of these, 33 (0.61%) were predicted to create a missense change, one a frameshift and another a stop gain, while 30 were synonymous (0.59%; Supplementary Table 5c). In total, 499 CCVs at 94 signals, and four additional HPPVs ($\geq 30\%$), are predicted to create new splice sites or activate cryptic splice sites in 126 genes (Supplementary Table 5d). These results are consistent

Table 1 | Signals with single CCVs and variants with a posterior probability of >80%

Fine-mapping region ^a	Variant ^b	Ref/alt ^c	EAF ^d	PP ^e	Model ^f	Signal ^g	CCVs ^h	ER negative		ER positive		P value ⁱ	FPI	Predicted target gene(s) ^k	Confidence ^l
								OR ^j	95% CI	OR ^j	95% CI				
Chr1:120723447-121780613	rs11249433	A/G	0.42	0.57	ER all	Signal1	1	1.02	0.99-1.04	1.13	1.11-1.15	8.11 × 10 ⁻⁶⁰	NA	NA	
Chr1:200937832-201937832	rs35383942	C/T	0.06	0.96	ER all	Signal1	2	1.10	1.05-1.16	1.09	1.06-1.13	1.14 × 10 ⁻⁷	D	TNNI7	Level 1
Chr2:201681247-202681247	rs3769821	C/T	0.66	0.40	ER all	Signal1	1	0.94	0.92-0.97	0.95	0.93-0.96	1.46 × 10 ⁻¹²	D	ALS2CR2	Level 1
Chr2:217405832-218796508	rs4442975 ^m	G/T	0.48	0.84	ER all	Signal1	1	0.94	0.92-0.97	0.86	0.85-0.87	2.50 × 10 ⁻⁹⁰	D	IGFBP5 ⁿ	Level 2
Chr4:105569013-106856761	esv3601665	-/Alu	0.07	0.95	ER pos			1.01	0.95-1.08	1.10	1.06-1.14	3.27 × 10 ⁻⁶	D	ARHGFB38 and AC004066.3	Level 1
Chr5:779790-1797488	rs10069690	C/T	0.27	0.58	ER neg	Signal1	1	1.18	1.15-1.21	1.03	1.01-1.05	1.20 × 10 ⁻³⁴	D	SIC6A18 and TER ^o	Level 2
Chr5:44013304-45206498	rs10941679	A/G	0.26	0.00	ER pos	Signal1	1	1.04	1.02-1.07	1.17	1.15-1.19	1.50 × 10 ⁻⁷⁷	D	MRF530	Level 2
	rs5867671	A/-	0.77	0.01	ER pos	Signal2	1	0.91	0.89-0.94	0.99	0.97-1.01	2.25 × 10 ⁻⁹	NA	NA	
Chr5:44013304-45206498	rs190443933	T/C	0.01	0.00	ER all	Signal4	1	1.30	1.14-1.48	1.26	1.16-1.37	2.32 × 10 ⁻⁸	NA	NA	
Chr5:55531884-56587883	rs984113	G/C	0.61	0.81	ER pos	Signal2	1	0.96	0.93-0.98	0.96	0.94-0.97	3.51 × 10 ⁻⁸	D	MAP3K7 ⁿ	Level 2
	rs889310	C/T	0.56	0.84	ER pos	(Signal 6)	15	1.03	1.00-1.05	1.05	1.03-1.06	1.75 × 10 ⁻⁷	D	MAP3K7 ⁿ	Level 1
Chr6:15899557-16899557	rs3819405	C/T	0.32	0.96	ER all	Signal1	1	0.97	0.95-1.00	0.95	0.94-0.97	1.14 × 10 ⁻⁷	D	ATXN1, RPT151F171 and RPT151F17.2	Level 2
Chr6:151418856-152937016	rs12173562	C/T	0.08	0.10	ER neg	Signal1	1	1.30	1.25-1.36	1.14	1.11-1.18	3.98 × 10 ⁻⁴⁰	D	ESR7 ⁿ	Level 1
	rs34133739	-/C	0.53	0.25	ER all	Signal2	1	1.11	1.09-1.14	1.05	1.04-1.07	2.36 × 10 ⁻²²	D	ESR7 ⁿ	Level 1
	rs851984	G/A	0.40	0.73	ER all	Signal3	1	1.07	1.04-1.09	1.05	1.04-1.07	3.69 × 10 ⁻¹³	D	ESR7 ⁿ	Level 1
Chr7:130167121-131167121	rs68056147	G/A	0.30	0.84	ER all			1.04	1.01-1.07	1.05	1.03-1.06	3.07 × 10 ⁻⁷	D	MKLM1	Level 2
Chr8:127424659-130041931	rs35961416	-/A	0.41	0.68	ER all	Signal3	1	0.97	0.94-0.99	0.95	0.93-0.96	9.97 × 10 ⁻¹¹	D	MYC ⁿ	Level 1
Chr9:21247803-22624477	rs539723051	AAA/-	0.33	0.43	ER all	Signal1	1	1.08	1.05-1.11	1.06	1.04-1.08	1.81 × 10 ⁻¹⁵	NA	NA	
Chr9:109803808-111395353	rs10816625	A/G	0.07	0.95	ER pos	Signal3	1	1.06	1.01-1.11	1.13	1.10-1.16	3.62 × 10 ⁻¹⁵	D	KLF4 ⁿ	Level 2
	rs13294895	C/T	0.18	0.93	ER pos	Signal4	1	1.01	0.98-1.05	1.09	1.07-1.11	4.00 × 10 ⁻¹⁷	D	KLF4 ⁿ	Level 1
Chr9:109803808-111395353	rs60037937	AA/-	0.22	0.68	ER pos	Signal2	1	1.02	0.99-1.06	1.11	1.09-1.13	3.17 × 10 ⁻²⁶	D	KLF4 ⁿ and RAD23B	Level 2
Chr10:63758684-65063702	rs10995201	A/G	0.15	0.31	ER all	Signal1	1	0.91	0.88-0.94	0.87	0.85-0.89	1.40 × 10 ⁻³⁷	NA	NA	
Chr10:122593901-123849324	rs35054928	C/-	0.56	0.60	ER all	Signal1	1	0.96	0.94-0.98	0.74	0.73-0.76	6.55 × 10 ⁻³⁴²	D	FGFR2 ⁿ	Level 1
	rs45631563 ^m	A/T	0.04	0.93	ER pos	Signal3	1	0.97	0.92-1.03	0.76	0.73-0.79	4.84 × 10 ⁻⁴⁴	C	FGFR2 ⁿ	Level 2
Chr11:68831418-69879161	rs7899765	T/C	0.06	0.02	ER all	Signal5	1	1.01	0.97-1.06	0.87	0.84-0.90	2.21 × 10 ⁻¹⁸	D	FGFR2 ⁿ	Level 1
Chr12:27639846-29034415	rs78540526	C/T	0.09	0.91	ER pos	Signal1	1	1.01	0.97-1.06	1.40	1.36-1.44	2.77 × 10 ⁻¹⁴⁵	D	CCND7 ⁿ and MYEOV	Level 1
Chr12:115336522-116336522	rs7297051	C/T	0.23	0.23	ER all	Signal1	1	0.87	0.85-0.90	0.89	0.88-0.91	3.12 × 10 ⁻⁴³	D	CCDC97 ⁿ , PTHLH ⁿ and RPT1-967K21.1	Level 2
Chr14:91341069-92368623	rs35422	G/A	0.57	0.58	ER pos	Signal2	1	0.98	0.96-1.01	1.05	1.03-1.07	4.85 × 10 ⁻¹⁰	D	TBX3	Level 1
	rs7153397	C/T	0.70	0.81	ER pos	Signal1	3	1.01	0.99-1.04	1.06	1.04-1.08	3.25 × 10 ⁻¹¹	D and C	CCDC88C, CTD-2547L24.4, C14orf159, GPR68, RPS6KA5, RPT1-73M18.7 and RPT1-895M11.3	Level 2
Chr16:52038825-53038825	rs4784227	C/T	0.27	0.95	ER pos	Signal1	1	1.15	1.12-1.18	1.26	1.24-1.28	4.63 × 10 ⁻¹⁶⁰	D	TOX3 ⁿ	Level 1
Chr18:23832476-25075396	rs180952292	T/C	0.01	0.01	ER neg	Signal4	1	1.24	1.12-1.37	0.98	0.92-1.05	2.07 × 10 ⁻⁵	NA	NA	
Chr18:41899590-42899590	rs9952980	T/C	0.34	0.95	ER all	Signal2	3	0.97	0.94-0.99	0.95	0.93-0.96	7.43 × 10 ⁻¹²	D	SIC14A2	Level 2
Chr20:5448227-6448227	rs16991615	G/A	0.07	0.97	ER all	Signal1	1	1.09	1.04-1.15	1.07	1.04-1.11	7.89 × 10 ⁻⁷	D and C	GPCPD1 and MCM8	Level 2
Chr22:45783297-46783297	rs184070480	C/T	0.01	0	ER all	Signal2	1	1.40	1.20-1.64	1.01	0.91-1.12	5.02 × 10 ⁻⁵	D	ATXN10 and WN178	Level 2

^aLabels show the chromosome (chr) number and range of base pairs (from genome assembly GRCh37/hg19). ^bCurrent reference identification. ^cReference (ref) versus alternative (alt) allele. ^dEffect (alt) allele frequency (EAF) in OncoArray. ^eLargest posterior probability (PP) in all of the evaluated models. ^fModel with which the variant reaches the largest PP (ER all; ER positive (ER pos) or ER negative (ER neg)). ^gSignal where the variant is included. Numbers in brackets represent moderate-confidence signals. ^hNumber of CCVs in the signal. ⁱMultinomial logistic regression summary statistics (odds ratios (ORs), confidence intervals (CIs) and χ^2 single-variant analysis P values), estimated using 67136 ER-positive and 17506 ER-negative cases, together with 88937 controls. ^jFunction prediction (FP) (distal regulation (D), proximal regulation (P), coding (C) or not available (N/A)). ^kPredicted target genes with the largest confidence level for each variant. ^lINQUISIT level of confidence. ^mTwo variants reach a PP of >0.8 in both the ER-all and ER-pos models. ⁿrs4442975 (ER-pos PP = 0.83; ER-all PP = 0.84); and rs45631563 (ER-pos PP = 0.93; ER-all PP = 0.92). ^oTarget genes with functional follow up.

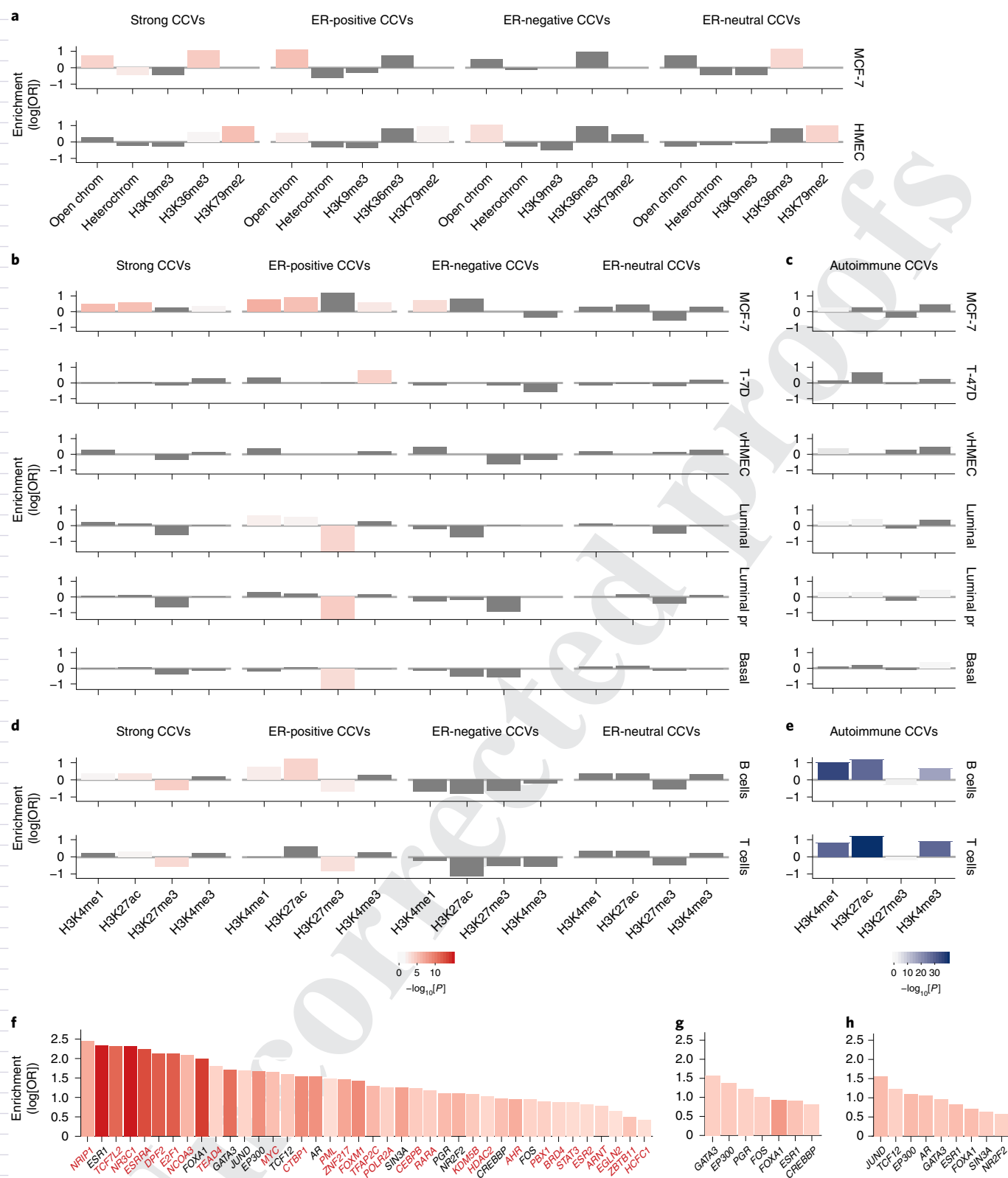


Fig. 3 | Overlap of CCVs with gene regulatory regions, gene bodies and TFBSs. a, Breast cancer CCVs overlap with chromatin states and broad breast cell epigenetic marks. HMEC, human mammary epithelial cells. **b, c**, Breast cancer CCVs (**b**) and autoimmune CCVs (**c**) overlap with breast cell epigenetic marks. vHMEC, variant HMEC. **d, e**, Breast cancer CCVs (**d**) autoimmune CCVs (**e**) and overlap with autoimmune-related epigenetic marks. In **a, b** and **d**, the column ‘strong CCVs’ represents analysis with all CCVs at strong signals, while the remaining columns represent analysis of CCVs at strong signals stratified by phenotype. Logistic regression robust variance estimation for clustered observations was used, and Wald test χ^2 P values were estimated using 67,136 ER-positive and 17,506 ER-negative cases, together with 88,937 controls. Non-significant P values are shown in dark gray. Significance was defined as an FDR of 5%, which corresponds to the following P value thresholds: $P=1.66 \times 10^{-2}$ (strong signals); $P=2.42 \times 10^{-2}$ (ER positive); $P=3.02 \times 10^{-3}$ (ER negative); and $P=1.76 \times 10^{-3}$ (ER neutral). **f–h**, Significant ER-positive (**f**), ER-negative (**g**) and ER-neutral CCVs (**h**) overlap with TFBSs. TFBSs found significant for ER-positive CCVs are highlighted in red (x axis labels).

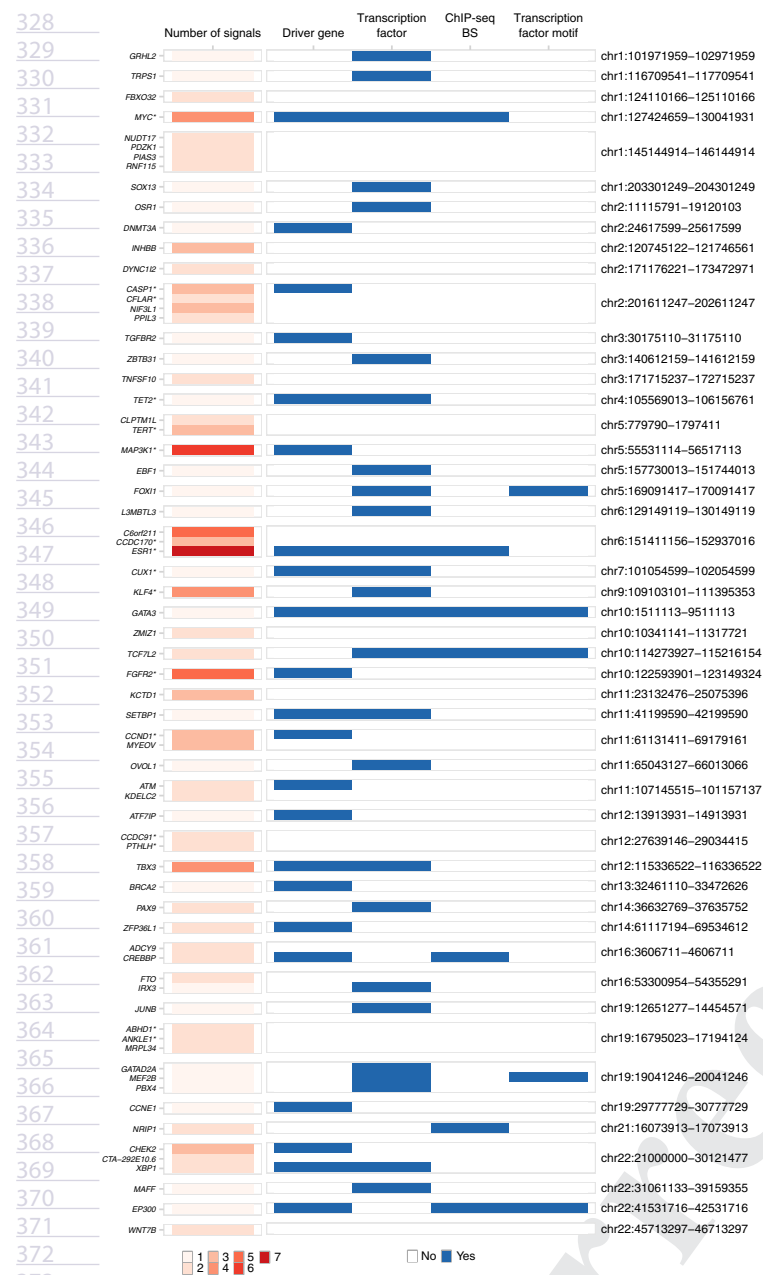


Fig. 4 | Predicted target genes are enriched in known breast cancer driver genes and transcription factors. Target genes ($n = 79$) that fulfill at least one of the following criteria: (1) is targeted by more than one independent signal; (2) is a known driver gene; (3) is a known transcription factor gene; (4) its binding sites (as determined by ChIP-Seq with bisulfite sequencing (ChIP-Seq)) are significantly overlapped by CCVs; or (5) its consensus (transcription factor) motif is significantly overlapped by CCVs. Asterisks indicate genes with published functional follow up.

with previous observations that the majority of common susceptibility variants are regulatory.

We applied an updated version of our pipeline INQUISIT (integrated expression quantitative trait and in silico prediction of GWAS targets)² to prioritize potential target genes from 5,375 CCVs in strong signals and all 138 HPPVs ($\geq 30\%$; Supplementary Table 2c). The pipeline predicted 1,204 target genes from 124 out of 128 genomic regions examined. As a validation, we examined the overlap between INQUISIT predictions and 278 established

breast cancer driver genes^{35–39}. Cancer driver genes were over-represented among high-confidence (level 1) targets, with a fivefold increase over expected levels from CCVs and a 15-fold increase from HPPVs ($P = 1 \times 10^{-6}$; Supplementary Fig. 4a). Notably, 13 cancer driver genes (*ATAD2*, *CASP8*, *CCND1*, *CHEK2*, *ESR1*, *FGFR2*, *GATA3*, *MAP3K1*, *MYC*, *SETBP1*, *TBX3*, *XBPI* and *ZFP36L1*) were predicted from the HPPVs derived from PAINTOR. Cancer driver gene status was consequently included as an additional weighting factor in the INQUISIT pipeline. Transcription factor genes⁴⁰ were also enriched among high-confidence targets predicted from both CCVs (twofold; $P = 4.6 \times 10^{-4}$) and HPPVs (2.5-fold; $P = 1.8 \times 10^{-2}$; Supplementary Fig. 4a).

In total, INQUISIT identified 191 target genes supported by strong evidence (Supplementary Table 6). Significantly more genes were targeted by multiple independent signals ($n = 165$) than expected by chance ($P = 4.3 \times 10^{-8}$; Supplementary Fig. 4b and Fig. 4). Six high-confidence predictions came only from HPPVs, although three of these (*IGFBP5*, *POMGNT1* and *WDYHVI*) had been predicted at lower confidence from CCVs. Target genes included 20 that were prioritized via potential coding/splicing changes (Supplementary Table 7), ten via promoter variants (Supplementary Table 8) and 180 via distal regulatory variants (Supplementary Table 9). We illustrate the genes prioritized via multiple lines of evidence in Fig. 4.

Three examples of INQUISIT using genomic features to identify predict target genes. Based on genome-wide chromosome conformation capture from human mammary epithelial cells (Hi-C) and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) data, *NRIP1* is a predicted target of intergenic CCVs and HPPVs at chr21q21 (Supplementary Fig. 5a). Multiple target genes were predicted at chr22q12, including the driver genes *CHEK2* and *XBPI* (Supplementary Fig. 5b). A third example at chr12q24.31 is a more complicated scenario with two level 1 targets: *RPLP0* (ref. ⁴¹) and a modulator of mammary progenitor cell expansion, *MSI1* (ref. ⁴²) (Supplementary Fig. 5c).

Target gene pathways include DNA integrity checkpoint, apoptosis and developmental processes and the immune system. We performed pathway analysis to identify common processes using INQUISIT high-confidence target protein-coding genes (Fig. 5a) and identified 488 Gene Ontology terms and 307 pathways at an FDR of 5% (Supplementary Table 10). These were grouped into 98 themes by common ancestor Gene Ontology terms, pathways or transcription factor classes (Fig. 5b). We found that 23% (14/60) of the ER-positive target genes were classified within developmental process pathways (including mammary development), 18% were classified in immune system pathways and a further 17% were classified in nuclear receptor pathways. Of the genes targeted by ER-neutral signals, 21% (18/87) were classified in developmental process pathways, 19% were classified in immune system pathways and a further 18% were classified in apoptotic process pathways. The top themes of genes targeted by ER-negative signals were DNA integrity checkpoint processes and the immune system, each of which contained 19% of genes (7/37), and apoptotic processes (16%).

Novel pathways revealed by this study include tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) signaling, the AP-2 transcription factors pathway, and regulation of I κ B kinase/nuclear factor- κ B (NF- κ B) signaling. Of note, the latter of these is specifically over-represented among ER-negative target genes. We also found significant over-representation of additional carcinogenesis-linked pathways, including cyclic adenosine monophosphate, NOTCH, phosphoinositide 3-kinase, RAS and WNT/ β -catenin, and of receptor tyrosine kinase signaling, including fibroblast growth factor receptor, epidermal growth factor receptor and transforming growth factor- β receptor^{43–47}. Finally, our target genes are also significantly over-represented in DNA damage checkpoint and

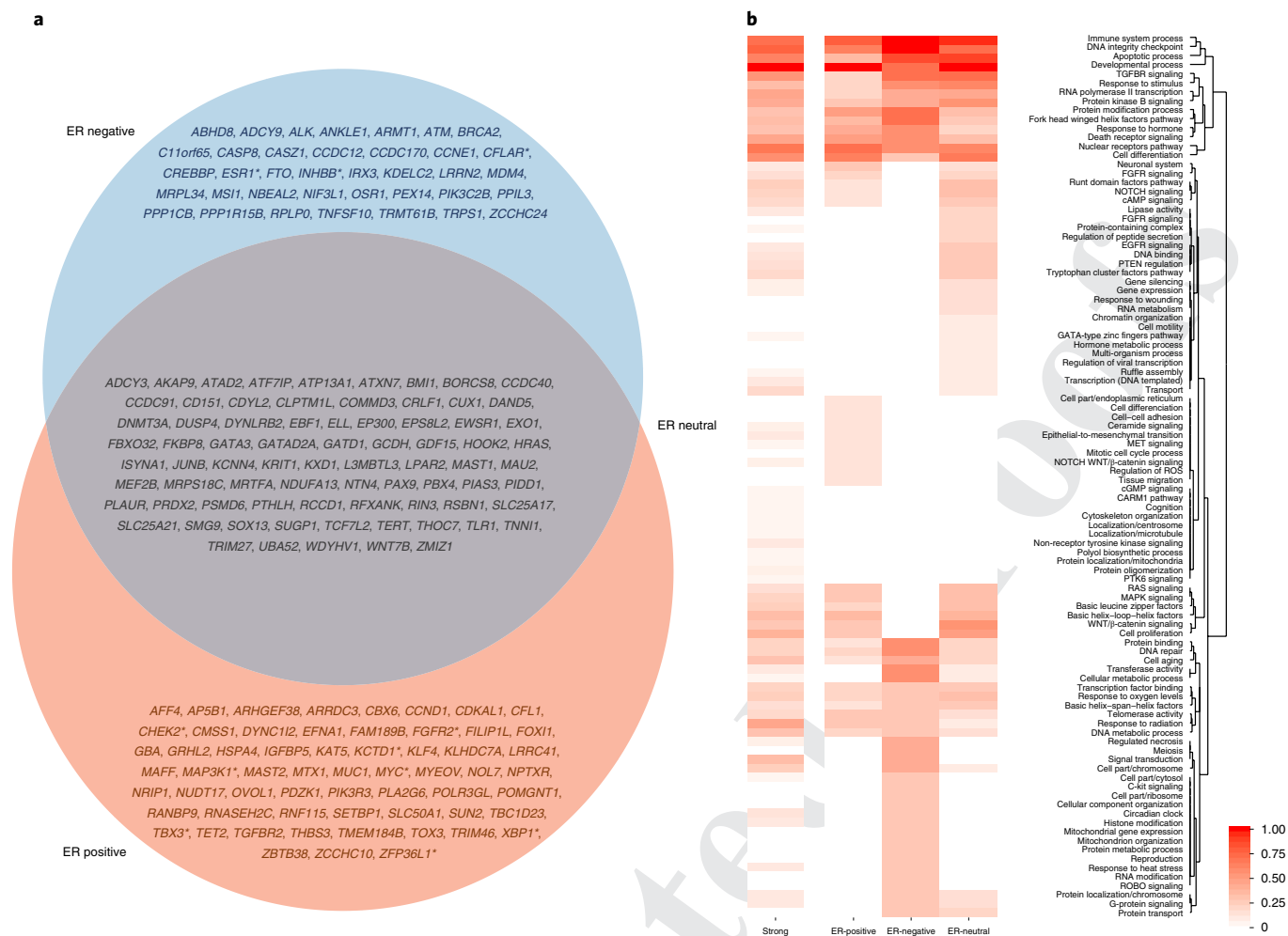


Fig. 5 | Predicted target genes by phenotype and significantly enriched pathways. a, Venn diagram showing the associated phenotype (ER positive, ER negative or ER neutral) for the level 1 target genes, predicted by the CCVs and HPPVs. Asterisks denote ER-positive or ER-negative target genes also targeted by ER-neutral signals. **b**, Heatmap showing clustering of pathway themes over-represented by INQUISIT level 1 target genes. Colors represent the relative number of genes per phenotype within enriched pathways, grouped by common themes (ER positive, ER negative, ER neutral or all phenotypes together (strong)).

394 DNA repair pathways, as well as programmed cell death pathways,
395 such as apoptotic processes, regulated necrosis and death receptor
396 signaling-related pathways.

398 Discussion

399 We have performed multiple, complementary analyses on 150
400 breast cancer-associated regions, originally found by GWASs, and
401 identified 362 independent risk signals, 205 of these with high confidence
402 ($P < 10^{-6}$). The inclusion of these new variants increases
403 the explained proportion of familial risk by 6% compared with that
404 explained by the lead signals alone.

405 We observed that most regions contain multiple independent
406 signals, with the greatest number (nine) in the region surrounding
407 *ESR1* and its co-regulated genes, and on 2q35, where *IGFBP5*
408 appears to be a key target. We used two complementary approaches
409 to identify likely causal variants within each region: a Bayesian
410 approach, PAINTOR (which integrated genetic associations, linkage
411 disequilibrium and informative genomic features, providing
412 complementary evidence) and a more traditional, multinomial
413 regression approach. PAINTOR supported most associations found
414 by multinomial regression and also identified additional variants.
415 Specifically, the Bayesian method highlighted 15 variants that are
416 highly likely to be causal (HPP $\geq 80\%$). From these approaches,

we identified a single variant, likely to be causal, at each of 34 signals
(Table 1). Of these, only rs16991615 (*MCM8*; NP_115874.3:p.E341K) and
rs7153397 (*CCDC88C*; NM_001080414.2:c.5058 + 134 2G > A; a cryptic splice-donor site) were predicted to affect protein-coding
sequences. However, in other signals, we also identified four coding changes previously recognized as deleterious: the stop gain
rs11571833 (*BRCA2*; NP_000050.2:p.K3326*)⁴⁸; two *CHEK2* coding variants (the frameshift rs555607708 (refs. 49,50) and a missense variant, rs17879961 (refs. 51,52)); and a splicing variant (rs10069690,
which in *TERT* results in the truncated protein INS1b¹⁹, decreased telomerase activity, telomere shortening and increased DNA damage response⁵³).

Having identified potential causal variants within each signal, we aimed to uncover their functions at the DNA level, as well as trying to predict their target gene(s). Looking across all 150 regions, a notable feature is that many likely causal variants implicated in ER-positive cancer risk lie in gene regulatory regions marked as open and active in ER-positive breast cells, but not in other cell types. Moreover, a significant proportion of potential causal variants overlap the binding sites for transcription factor proteins ($n = 40$ from ChIP-Seq) and co-regulators ($n = 64$ with the addition of computationally derived motifs). Furthermore, nine proteins also appear in the list of high-confidence target genes; hence, the following

417 genes and their products have been implicated by two different
418 approaches: *CREBBP*, *EP300*, *ESR1*, *FOXI1*, *GATA3*, *MEF2B*, *MYC*,
419 *NRIP1* and *TCF7L2*. Most proteins encoded by these genes already
420 have established roles in estrogen signaling. *CREBBP*, *EP300*, *ESR1*,
421 *GATA3* and *MYC* are also known cancer driver genes that are fre-
422 quently somatically mutated in breast tumors.

423 In contrast with ER-positive signals, we identified fewer genomic
424 features enriched in ER-negative signals. This may reflect the com-
425 mon molecular mechanisms underlying their development, but the
426 power of this study was limited, despite including as many patients
427 with ER-negative tumors as possible from the BCAC and CIMBA
428 consortia. Less than 20% of genomic signals confer a greater risk of
429 ER-negative cancer and there are few publicly available ChIP-Seq
430 data on ER-negative breast cancer cell lines. The heterogeneity of
431 ER-negative tumors also may have limited our power. Nevertheless,
432 we have identified 35 target genes for ER-negative likely causal vari-
433 ants. Some of these already had functional evidence supporting
434 their role: including *CASP8* (ref. ⁵⁴) and *MDM4* (ref. ⁵⁵). However,
435 most targets currently have no reported function in ER-negative
436 breast cancer development.

437 Finally, we examined the Gene Ontology pathways in which tar-
438 get genes most often lie. Of note, 14% (25/180) of all high-confi-
439 dence target genes and 19% of ER-negative target predictions are in
440 immune system pathways. Among the significantly enriched path-
441 ways were T cell activation, interleukin signaling, Toll-like receptor
442 cascades and I- κ B kinase/NF- κ B signaling, as well as processes lead-
443 ing to activation and perpetuation of the innate immune system.
444 The link between immunity, inflammation and tumorigenesis has
445 been studied extensively⁵⁶, although not primarily in the context of
446 susceptibility. Five ER-negative high-confidence target genes (*ALK*,
447 *CASP8*, *CFLAR*, *ESR1* and *TNFSF10*) lie in the I- κ B kinase/NF- κ B
448 signaling pathway. Interestingly, ER-negative cells have high lev-
449 els of NF- κ B activity compared with ER-positive cells⁵⁷. A recent
450 expression–methylation analysis on breast cancer tumor tissue also
451 identified clusters of genes correlated with DNA methylation levels:
452 one enriched in ER signaling genes and a second in immune path-
453 way genes⁵⁸.

454 These analyses provide strong evidence for more than 200 inde-
455 pendent breast cancer risk signals, identify the plausible cancer
456 variants and define likely target genes for the majority of these.
457 However, notwithstanding the enrichment of certain pathways and
458 transcription factors, the biological basis underlying most of these
459 signals remains poorly understood. Our analyses provide a rational
460 basis for such future studies into the biology underlying breast can-
461 cer susceptibility.

463 Online content

464 Any methods, additional references, Nature Research reporting
465 summaries, source data, extended data, supplementary informa-
466 tion, acknowledgements, peer review information; details of author
467 contributions and competing interests; and statements of data and
468 code availability are available at <https://doi.org/10.1038/s41588-019-0537-1>.

471 Received: 13 December 2018; Accepted: 24 October 2019;

474 References

- 475 1. Milne, R. L. et al. Identification of ten variants associated with risk of
476 estrogen-receptor-negative breast cancer. *Nat. Genet.* **49**, 1767–1778 (2017).
- 477 2. Michailidou, K. et al. Association analysis identifies 65 new breast cancer
478 risk loci. *Nature* **551**, 92–94 (2017).
- 479 3. Ghoussaini, M. et al. Evidence that breast cancer risk at the 2q35 locus is
480 mediated through IGFBP5 regulation. *Nat. Commun.* **4**, 4999 (2014).
- 481 4. Wyszynski, A. et al. An intergenic risk locus containing an enhancer
482 deletion in 2q35 modulates breast cancer risk by deregulating IGFBP5
expression. *Hum. Mol. Genet.* **25**, 3863–3876 (2016).
5. Guo, X. et al. Fine-scale mapping of the 4q24 locus identifies two
independent loci associated with breast cancer risk. *Cancer Epidemiol.
Biomark. Prev.* **24**, 1680–1691 (2015).
6. Glubb, D. M. et al. Fine-scale mapping of the 5q11.2 breast cancer locus
reveals at least three independent risk variants regulating MAP3K1. *Am. J.
Hum. Genet.* **96**, 5–20 (2015).
7. Dunning, A. M. et al. Breast cancer risk variants at 6q25 display different
phenotype associations and regulate *ESR1*, *RMND1* and *CCDC170*. *Nat.
Genet.* **48**, 374–386 (2016).
8. Shi, J. et al. Fine-scale mapping of 8q24 locus identifies multiple independent
risk variants for breast cancer. *Int. J. Cancer* **139**, 1303–1317 (2016).
9. Orr, N. et al. Fine-mapping identifies two additional breast cancer
susceptibility loci at 9q31.2. *Hum. Mol. Genet.* **24**, 2966–2984 (2015).
10. Darabi, H. et al. Polymorphisms in a putative enhancer at the 10q21.2
breast cancer risk locus regulate *NRBF2* expression. *Am. J. Hum. Genet.* **97**,
22–34 (2015).
11. Darabi, H. et al. Fine scale mapping of the 17q22 breast cancer locus using
dense SNPs, genotyped within the Collaborative Oncological Gene-
Environment Study (COGS). *Sci. Rep.* **6**, 32512 (2016).
12. Meyer, K. B. et al. Fine-scale mapping of the FGFR2 breast cancer risk
locus: putative functional variants differentially bind FOXA1 and E2F1. *Am.
J. Hum. Genet.* **93**, 1046–1060 (2013).
13. Betts, J. A. et al. Long noncoding RNAs *CUPID1* and *CUPID2* mediate
breast cancer risk at 11q13 by modulating the response to DNA damage.
Am. J. Hum. Genet. **101**, 255–266 (2017).
14. French, J. D. et al. Functional variants at the 11q13 risk locus for breast
cancer regulate cyclin D1 expression through long-range enhancers. *Am. J.
Hum. Genet.* **92**, 489–503 (2013).
15. Ghoussaini, M. et al. Evidence that the 5p12 variant rs10941679 confers
susceptibility to estrogen-receptor-positive breast cancer through *FGF10*
and *MRPS30* regulation. *Am. J. Hum. Genet.* **99**, 903–911 (2016).
16. Horné, H. N. et al. Fine-mapping of the 1p11.2 breast cancer susceptibility
locus. *PLoS ONE* **11**, e0160316 (2016).
17. Zeng, C. et al. Identification of independent association signals and putative
functional variants for breast cancer risk through fine-scale mapping of the
12p11 locus. *Breast Cancer Res.* **18**, 64 (2016).
18. Lin, W. Y. et al. Identification and characterization of novel associations in
the *CASP8/ALS2CR12* region on chromosome 2 with breast cancer risk.
Hum. Mol. Genet. **24**, 285–298 (2015).
19. Bojesen, S. E. et al. Multiple independent variants at the *TERT* locus are
associated with telomere length and risks of breast and ovarian cancer. *Nat.
Genet.* **45**, 371–384.e2 (2013).
20. Lawrenson, K. et al. Functional mechanisms underlying pleiotropic risk
alleles at the 19p13.1 breast–ovarian cancer susceptibility locus. *Nat.
Commun.* **7**, 12675 (2016).
21. Amos, C. I. et al. The OncoArray Consortium: a network for understanding
the genetic architecture of common cancers. *Cancer Epidemiol. Biomark.
Prev.* **26**, 126–135 (2017).
22. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci
associated with breast cancer risk. *Nat. Genet.* **45**, 353–361.e2 (2013).
23. Michailidou, K. et al. Genome-wide association analysis of more than
120,000 individuals identifies 15 new susceptibility loci for breast cancer.
Nat. Genet. **47**, 373–380 (2015).
24. Udler, M. S., Tyrer, J. & Easton, D. F. Evaluating the power to discriminate
between highly correlated SNPs in genetic association studies. *Genet.
Epidemiol.* **34**, 463–468 (2010).
25. Mavaddat, N., Antoniou, A. C., Easton, D. F. & Garcia-Closas, M. Genetic
susceptibility to breast cancer. *Mol. Oncol.* **4**, 174–191 (2010).
26. Lakhani, S. R. et al. Prediction of *BRCA1* status in patients with breast
cancer using estrogen receptor and basal phenotype. *Clin. Cancer Res.* **11**,
5175–5180 (2005).
27. Taberlay, P. C., Statham, A. L., Kelly, T. K., Clark, S. J. & Jones, P. A.
Reconfiguration of nucleosome-depleted regions at distal regulatory
elements accompanies DNA methylation of enhancers and insulators in
cancer. *Genome Res.* **24**, 1421–1432 (2014).
28. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease.
Cell **155**, 934–947 (2013).
29. Farh, K. K. et al. Genetic and epigenetic fine mapping of causal
autoimmune disease variants. *Nature* **518**, 337–343 (2015).
30. Cowper-Salari, R. et al. Breast cancer risk-associated SNPs modulate the
affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.* **44**,
1191–1198 (2012).
31. Kichaev, G. et al. Integrating functional data to prioritize causal variants in
statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
32. Quiroz-Zarate, A. et al. Expression quantitative trait loci (QTL) in tumor
adjacent normal breast tissue and breast tumor tissue. *PLoS ONE* **12**,
e0170181 (2017).
33. Cancer Genome Atlas Research Network et al. The cancer genome atlas
pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).

- 483 34. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast
484 tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- 485 35. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular
486 breast cancer. *Cell* **163**, 506–519 (2015).
- 487 36. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer
488 whole-genome sequences. *Nature* **534**, 47–54 (2016).
- 489 37. Pereira, B. et al. The somatic mutation profiles of 2,433 breast cancers
490 refines their genomic and transcriptomic landscapes. *Nat. Commun.* **7**,
491 11479 (2016).
- 492 38. Cancer Genome Atlas Network Comprehensive molecular portraits of
493 human breast tumours. *Nature* **490**, 61–70 (2012).
- 494 39. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes
495 and mutations. *Cell* **173**, 371–385.e18 (2018).
- 496 40. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665
497 (2018).
- 498 41. Artero-Castro, A. et al. Disruption of the ribosomal P complex leads to
499 stress-induced autophagy. *Autophagy* **11**, 1499–1519 (2015).
- 500 42. Wang, X. Y. et al. Musashi1 modulates mammary progenitor cell expansion
501 through proliferin-mediated activation of the Wnt and Notch pathways.
502 *Mol. Cell Biol.* **28**, 3589–3599 (2008).
- 503 43. Vijayan, D., Young, A., Teng, M. W. L. & Smyth, M. J. Targeting
504 immunosuppressive adenosine in cancer. *Nat. Rev. Cancer* **17**, 709–724 (2017).
- 505 44. Takebe, N. et al. Targeting Notch, Hedgehog, and Wnt pathways in cancer
506 stem cells: clinical update. *Nat. Rev. Clin. Oncol.* **12**, 445–464 (2015).
- 507 45. Thorpe, L. M., Yuzugullu, H. & Zhao, J. J. PI3K in cancer: divergent roles of
508 isoforms, modes of activation and therapeutic targeting. *Nat. Rev. Cancer*
509 **15**, 7–24 (2015).
- 510 46. Nusse, R. & Clevers, H. Wnt/ β -catenin signaling, disease, and emerging
511 therapeutic modalities. *Cell* **169**, 985–999 (2017).
- 512 47. Massague, J. TGF β signalling in context. *Nat. Rev. Mol. Cell Biol.* **13**,
513 616–630 (2012).
- 514 48. Meeks, H. D. et al. BRCA2 polymorphic stop codon K3326X and the risk of
515 breast, prostate, and ovarian cancers. *J. Natl. Cancer Inst.* **108**, djv315 (2016).
- 516 49. CHEK2 Breast Cancer Case-Control Consortium CHEK2*1100delC and
517 susceptibility to breast cancer: a collaborative analysis involving 10,860
518 breast cancer cases and 9,065 controls from 10 studies. *Am. J. Hum. Genet.*
519 **74**, 1175–1182 (2004).
- 520 50. Schmidt, M. K. et al. Age- and tumor subtype-specific breast cancer risk
521 estimates for CHEK2*1100delC carriers. *J. Clin. Oncol.* **34**, 2750–2760
522 (2016).
- 523 51. Kilpivaara, O. et al. CHEK2 variant I157T may be associated with increased
524 breast cancer risk. *Int. J. Cancer* **111**, 543–547 (2004).
- 525 52. Muranen, T. A. et al. Patient survival and tumor characteristics associated
526 with CHEK2:p.I157T—findings from the Breast Cancer Association
527 Consortium. *Breast Cancer Res.* **18**, 98 (2016).
- 528 53. Killedar, A. et al. A common cancer risk-associated allele in the *hTERT*
529 locus encodes a dominant negative inhibitor of telomerase. *PLoS Genet.* **11**,
530 e1005286 (2015).
- 531 54. De Basio, A. et al. Unusual roles of caspase-8 in triple-negative breast
532 cancer cell line MDA-MB-231. *Int. J. Oncol.* **48**, 2339–2348 (2016).
- 533 55. Haupt, S. et al. Targeting Mdmx to treat breast cancers with wild-type p53.
534 *Cell Death Dis.* **6**, e1821 (2015).
- 535 56. Pandya, P. H., Murray, M. E., Pollok, K. E. & Renbarger, J. L. The immune
536 system in cancer pathogenesis: potential therapeutic approaches. *J.*
537 *Immunol. Res.* **2016**, 4273943 (2016).
- 538 57. Gionet, N., Jansson, D., Mader, S. & Pratt, M. A. NF- κ B and estrogen
539 receptor α interactions: differential function in estrogen receptor-negative
540 and -positive hormone-independent breast cancer cells. *J. Cell Biochem.*
541 **107**, 448–459 (2009).
- 542 58. Fleischer, T. et al. DNA methylation at enhancers identifies distinct breast
543 cancer lineages. *Nat. Commun.* **8**, 1379 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Laura Fachal¹, Hugues Aschard^{2,3,4,458}, Jonathan Beesley^{5,458}, Daniel R. Barnes⁶, Jamie Allen⁶,
Siddhartha Kar¹, Karen A. Pooley⁶, Joe Dennis⁶, Kyriaki Michailidou^{6,7}, Constance Turman⁴,
Penny Soucy⁸, Audrey Lemaçon⁸, Michael Lush⁶, Jonathan P. Tyrer¹, Maya Ghoussaini¹,
Mahdi Moradi Marjaneh⁵, Xia Jiang³, Simona Agata⁹, Kristiina Aittomäki¹⁰, M. Rosario Alonso¹¹,
Irene L. Andrulis^{12,13}, Hoda Anton-Culver¹⁴, Natalia N. Antonenkova¹⁵, Adalgeir Arason^{16,17},
Volker Arndt¹⁸, Kristan J. Aronson¹⁹, Banu K. Arun²⁰, Bernd Auber²¹, Paul L. Auer^{22,23},
Jacopo Azzollini²⁴, Judith Balmaña^{25,26}, Rosa B. Barkardottir^{16,17}, Daniel Barrowdale⁶,
Alicia Beeghly-Fadiel²⁷, Javier Benitez^{28,29}, Marina Bermisheva³⁰, Katarzyna Białkowska³¹,
Amie M. Blanco³², Carl Blomqvist^{33,34}, William Blot^{27,35}, Natalia V. Bogdanova^{15,36,37},
Stig E. Bojesen^{38,39,40}, Manjeet K. Bolla⁶, Bernardo Bonanni⁴¹, Ake Borg⁴², Kristin Bosse⁴³,
Hiltrud Brauch^{44,45,46}, Hermann Brenner^{18,46,47}, Ignacio Briceno^{48,49}, Ian W. Brock⁵⁰,
Angela Brooks-Wilson^{51,52}, Thomas Brüning⁵³, Barbara Burwinkel^{54,55}, Sandra S. Buys⁵⁶, Qiuyin Cai²⁷,
Trinidad Caldés⁵⁷, Maria A. Caligo⁵⁸, Nicola J. Camp⁵⁹, Ian Campbell^{60,61}, Federico Canzian⁶²,
Jason S. Carroll⁶³, Brian D. Carter⁶⁴, Jose E. Castela⁶⁵, Jocelyne Chiquette⁶⁶, Hans Christiansen³⁶,
Wendy K. Chung⁶⁷, Kathleen B. M. Claes⁶⁸, Christine L. Clarke⁶⁹, GEMO Study Collaborators⁴⁶⁰,
EMBRACE Collaborators⁴⁶⁰, J. Margriet Collée⁷⁰, Sten Cornelissen⁷¹, Fergus J. Couch⁷²,
Angela Cox⁵⁰, Simon S. Cross⁷³, Cezary Cybulski³¹, Kamila Czene⁷⁴, Mary B. Daly⁷⁵,
Miguel de la Hoya⁵⁷, Peter Devilee^{76,77}, Orland Diez^{78,79}, Yuan Chun Ding⁸⁰, Gillian S. Dite⁸¹,
Susan M. Domchek⁸², Thilo Dörk³⁷, Isabel dos-Santos-Silva⁸³, Arnaud Droit^{8,84}, Stéphane Dubois⁸,
Martine Dumont⁸, Mercedes Duran⁸⁵, Lorraine Durcan^{86,87}, Miriam Dwek⁸⁸, Diana M. Eccles⁸⁹,
Christoph Engel⁹⁰, Mikael Eriksson⁷⁴, D. Gareth Evans^{91,92}, Peter A. Fasching^{93,94}, Olivia Fletcher⁹⁵,
Giuseppe Floris⁹⁶, Henrik Flyger⁹⁷, Lenka Foretova⁹⁸, William D. Foulkes⁹⁹, Eitan Friedman^{100,101},
Lin Fritschi¹⁰², Debra Frost⁶, Marike Gabrielson⁷⁴, Manuela Gago-Dominguez^{103,104},

Gaetana **Gambino**⁵⁸, Patricia A. **Ganz**¹⁰⁵, Susan M. **Gapstur**⁶⁴, Judy **Garber**¹⁰⁶,
 José A. **García-Sáenz** ¹⁰⁷, Mia M. **Gaudet**⁶⁴, Vassilios **Georgoulis**¹⁰⁸, Graham G. **Giles**^{81,109,110},
 Gord **Glendon**¹², Andrew K. **Godwin**¹¹¹, Mark S. **Goldberg**^{112,113}, David E. **Goldgar**¹¹⁴,
 Anna **González-Neira**²⁹, Maria Grazia **Tibiletti**¹¹⁵, Mark H. **Greene** ¹¹⁶, Mervi **Grip**¹¹⁷, Jacek **Gronwald**³¹,
 Anne **Grundy**¹¹⁸, Pascal **Guénel** ¹¹⁹, Eric **Hahnen**^{120,121}, Christopher A. **Haiman**¹²², Niclas **Håkansson**¹²³,
 Per **Hall**^{74,124}, Ute **Hamann**¹²⁵, Patricia A. **Harrington**¹, Jaana M. **Hartikainen**^{126,127,128},
 Mikael **Hartman**^{129,130}, Wei **He**⁷⁴, Catherine S. **Healey**¹, Bernadette A. M. **Heemskerk-Gerritsen** ¹³¹,
 Jane **Heyworth**¹³², Peter **Hillemanns**³⁷, Frans B. L. **Hogervorst**¹³³, Antoinette **Hollestelle** ¹³¹,
 Maartje J. **Hooning**¹³¹, John L. **Hopper**⁸¹, Anthony **Howell**¹³⁴, Guanmengqian **Huang**¹²⁵,
 Peter J. **Hulick** ^{135,136}, Evgeny N. **Imyanitov**¹³⁷, KConFab Investigators⁴⁶⁰, HEBON Investigators⁴⁶⁰,
 ABCTB Investigators, Claudine **Isaacs**¹³⁸, Motoki **Iwasaki**¹³⁹, Agnes **Jager**¹³¹, Milena **Jakimovska** ¹⁴⁰,
 Anna **Jakubowska** ^{31,141}, Paul A. **James** ^{61,142}, Ramunas **Janavicius**^{143,144}, Rachel C. **Jankowitz**¹⁴⁵,
 Esther M. **John**¹⁴⁶, Nichola **Johnson**⁹⁵, Michael E. **Jones** ¹⁴⁷, Arja **Jukkola-Vuorinen**¹⁴⁸, Audrey **Jung**¹⁴⁹,
 Rudolf **Kaaks**¹⁴⁹, Daehee **Kang**^{150,151,152}, Pooja Middha **Kapoor** ^{149,153}, Beth Y. **Karlan**^{154,155},
 Renske **Keeman** ⁷¹, Michael J. **Kerin**¹⁵⁶, Elza **Khusnutdinova**^{30,157}, Johanna I. **Kiiski**¹⁵⁸, Judy **Kirk**¹⁵⁹,
 Cari M. **Kitahara**¹⁶⁰, Yon-Dschun **Ko**¹⁶¹, Irene **Konstantopoulou** ¹⁶², Veli-Matti **Kosma**^{126,127,128},
 Stella **Koutros**¹⁶³, Katerina **Kubelka-Sabit**¹⁶⁴, Ava **Kwong**^{165,166,167}, Kyriacos **Kyriacou**⁷, Yael **Laitman**¹⁰⁰,
 Diether **Lambrechts**^{168,169}, Eunjung **Lee**¹²², Goska **Leslie** ⁶, Jenny **Lester**^{154,155},
 Fabienne **Lesueur** ^{170,171,172}, Annika **Lindblom**^{173,174}, Wing-Yee **Lo**^{44,45}, Jirong **Long**²⁷,
 Artitaya **Lophatananon**^{175,176}, Jennifer T. **Loud**¹¹⁶, Jan **Lubiński**³¹, Robert J. **MacInnis**^{81,109},
 Tom **Maishman**^{86,87}, Enes **Makalic**⁸¹, Arto **Mannermaa**^{126,127,128}, Mehdi **Manoochehri**¹²⁵,
 Siranoush **Manoukian**²⁴, Sara **Margolin**^{124,177}, Maria Elena **Martinez**^{104,178}, Keitaro **Matsuo** ^{179,180},
 Tabea **Maurer**¹⁸¹, Dimitrios **Mavroudis**¹⁰⁸, Rebecca **Mayes**¹, Lesley **McGuffog**⁶, Catriona **McLean** ¹⁸²,
 Noura **Mebirouk**^{170,171,183}, Alfons **Meindl**¹⁸⁴, Austin **Miller** ¹⁸⁵, Nicola **Miller**¹⁵⁶, Marco **Montagna**⁹,
 Fernando **Moreno**¹⁰⁷, Kenneth **Muir** ^{175,176}, Anna Marie **Mulligan**^{186,187}, Victor M. **Muñoz-Garzon**¹⁸⁸,
 Taru A. **Muranen** ¹⁵⁸, Steven A. **Narod**¹⁸⁹, Rami **Nassir**¹⁹⁰, Katherine L. **Nathanson** ⁸²,
 Susan L. **Neuhausen**⁸⁰, Heli **Nevanlinna** ¹⁵⁸, Patrick **Neven**⁹⁶, Finn C. **Nielsen**¹⁹¹, Liene **Nikitina-Zake**¹⁹²,
 Aaron **Norman**¹⁹³, Kenneth **Offit**^{194,195}, Edith **Olah**¹⁹⁶, Olufunmilayo I. **Olopade** ¹⁹⁷, Håkan **Olsson**¹⁹⁸,
 Nick **Orr** ¹⁹⁹, Ana **Osorio**^{28,29}, V. Shane **Pankratz** ²⁰⁰, Janos **Papp**¹⁹⁶, Sue K. **Park**^{150,151,152},
 Tjong-Won **Park-Simon**³⁷, Michael T. **Parsons** ⁵, James **Paul** ²⁰¹, Inge Sokilde **Pedersen**^{202,203,204},
 Bernard **Peissel**²⁴, Beth **Peshkin** ¹³⁸, Paolo **Peterlongo** ²⁰⁵, Julian **Peto** ⁸³,
 Dijana **Plaseska-Karanfilska** ¹⁴⁰, Karolina **Prajzencanc**³¹, Ross **Prentice**²², Nadege **Presneau**⁸⁸,
 Darya **Prokofyeva**¹⁵⁷, Miquel Angel **Pujana**²⁰⁶, Katri **Pyhkäs** ^{207,208}, Paolo **Radice** ²⁰⁹,
 Susan J. **Ramus**^{210,211}, Johanna **Rantala**²¹², Rohini **Rau-Murthy**¹⁹⁵, Gad **Rennert** ²¹³, Harvey A. **Risch**²¹⁴,
 Mark **Robson** ¹⁹⁵, Atocha **Romero** ²¹⁵, Caroline Maria **Rossing**¹⁹¹, Emmanouil **Saloustros** ²¹⁶,
 Estela **Sánchez-Herrero**²¹⁵, Dale P. **Sandler**²¹⁷, Marta **Santamariña**^{28,218,219}, Christobel **Saunders**²²⁰,
 Elinor J. **Sawyer**²²¹, Maren T. **Scheuner**³², Daniel F. **Schmidt**^{81,222}, Rita K. **Schmutzler**^{120,121},
 Andreas **Schneeweiss**^{55,223}, Minouk J. **Schoemaker**¹⁴⁷, Ben **Schöttker**^{18,224}, Peter **Schürmann**³⁷,
 Christopher **Scott** ¹⁹³, Rodney J. **Scott** ^{225,226,227}, Leigha **Senter**²²⁸, Caroline M. **Seynaeve**¹³¹,
 Mitul **Shah**¹, Priyanka **Sharma**²²⁹, Chen-Yang **Shen**^{230,231}, Xiao-Ou **Shu**²⁷, Christian F. **Singer**²³²,
 Thomas P. **Slavin**²³³, Snezhana **Smichkoska**²³⁴, Melissa C. **Southey**^{110,235}, John J. **Spinelli**^{236,237},
 Amanda B. **Spurdle** ⁵, Jennifer **Stone** ^{81,238}, Dominique **Stoppa-Lyonnet**^{183,239,240},
 Christian **Sutter** ²⁴¹, Anthony J. **Swerdlow**^{147,242}, Rulla M. **Tamimi**^{3,4,243}, Yen Yen **Tan** ²⁴⁴,

William J. Tapper⁸⁹, Jack A. Taylor^{217,245}, Manuel R. Teixeira^{246,247}, Maria Tengström^{126,248,249}, Soo H. Teo^{250,251}, Mary Beth Terry²⁵², Alex Teulé²⁵³, Mads Thomassen²⁵⁴, Darcy L. Thull²⁵⁵, Marc Tischkowitz^{99,256}, Amanda E. Toland²⁵⁷, Rob A. E. M. Tollenaar²⁵⁸, Ian Tomlinson^{259,260}, Diana Torres^{48,125}, Gabriela Torres-Mejía²⁶¹, Melissa A. Troester²⁶², Thérèse Truong¹¹⁹, Nadine Tung²⁶³, Maria Tzardi²⁶⁴, Hans-Ulrich Ulmer²⁶⁵, Celine M. Vachon²⁶⁶, Christi J. van Asperen²⁶⁷, Lizet E. van der Kolk¹³³, Elizabeth J. van Rensburg²⁶⁸, Ana Vega²⁶⁹, Alessandra Viel²⁷⁰, Joseph Vijai^{194,195}, Maartje J. Vogel¹³³, Qin Wang⁶, Barbara Wappenschmidt^{120,121}, Clarice R. Weinberg²⁷¹, Jeffrey N. Weitzel²³³, Camilla Wendt¹⁷⁷, Hans Wildiers⁹⁶, Robert Winqvist^{207,208}, Alicja Wolk^{123,272}, Anna H. Wu¹²², Drakoulis Yannoukakos¹⁶², Yan Zhang^{18,46}, Wei Zheng²⁷, David Hunter^{3,4}, Paul D. P. Pharoah^{1,6}, Jenny Chang-Claude^{149,181}, Montserrat Garcia-Closas^{163,273}, Marjanka K. Schmidt^{71,274}, Roger L. Milne^{81,109,110}, Vessela N. Kristensen^{275,276,277,278}, Juliet D. French⁵, Stacey L. Edwards⁵, Antonis C. Antoniou⁶, Georgia Chenevix-Trench^{5,459}, Jacques Simard^{8,459}, Douglas F. Easton^{1,6,459}, Peter Kraft^{3,4,459*} and Alison M. Dunning^{1,459*}

¹Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. ²Centre de Bioinformatique Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France. ³Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁴Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁵Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ⁶Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁷Department of Electron Microscopy/Molecular Pathology and The Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. ⁸Genomics Center, Centre de Recherche, Centre Hospitalier Universitaire de Québec, Université Laval, Québec City, Québec, Canada. ⁹Immunology and Molecular Oncology Unit, Veneto Institute of Oncology (IOV), IRCCS, Padua, Italy. ¹⁰Department of Clinical Genetics, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. ¹¹Human Genotyping-CEGEN Unit, Human Cancer Genetic Program, Spanish National Cancer Research Centre, Madrid, Spain. ¹²Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. ¹³Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ¹⁴Department of Epidemiology, Genetic Epidemiology Research Institute, University of California, Irvine, Irvine, CA, USA. ¹⁵N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus. ¹⁶Department of Pathology, Landspítali University Hospital, Reykjavik, Iceland. ¹⁷BMC (Biomedical Centre), Faculty of Medicine, University of Iceland, Reykjavik, Iceland. ¹⁸Division of Clinical Epidemiology and Aging Research (C070), German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁹Department of Public Health Sciences and Cancer Research Institute, Queen's University, Kingston, Ontario, Canada. ²⁰Department of Breast Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA. ²¹Institute of Human Genetics, Hannover Medical School, Hannover, Germany. ²²Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ²³Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA. ²⁴Unit of Medical Genetics, Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy. ²⁵High Risk and Cancer Prevention Group, Vall Hebron Institute of Oncology, Barcelona, Spain. ²⁶Department of Medical Oncology, Vall Hebron University Hospital, Barcelona, Spain. ²⁷Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA. ²⁸Centro de Investigación en Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain. ²⁹Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. ³⁰Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia. ³¹Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. ³²Cancer Genetics and Prevention Program, University of California, San Francisco, San Francisco, CA, USA. ³³Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. ³⁴Department of Oncology, Örebro University Hospital, Örebro, Sweden. ³⁵International Epidemiology Institute, Rockville, MD, USA. ³⁶Department of Radiation Oncology, Hannover Medical School, Hannover, Germany. ³⁷Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. ³⁸Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. ³⁹Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. ⁴⁰Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ⁴¹Division of Cancer Prevention and Genetics, European Institute of Oncology (IEO), IRCCS, Milan, Italy. ⁴²Department of Oncology, Lund University and Skåne University Hospital, Lund, Sweden. ⁴³Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany. ⁴⁴Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany. ⁴⁵iFIT Cluster of Excellence, University of Tuebingen, Tuebingen, Germany. ⁴⁶German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴⁷Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. ⁴⁸Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota, Colombia. ⁴⁹Medical Faculty, Universidad de La Sabana, Bogota, Colombia. ⁵⁰Sheffield Institute for Nucleic Acids (SInFoNIA), Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK. ⁵¹Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada. ⁵²Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada. ⁵³Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany. ⁵⁴Molecular Epidemiology Group (C080), German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁵⁵Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Germany. ⁵⁶Department of Medicine, Huntsman Cancer Institute, Salt Lake City, UT, USA. ⁵⁷Molecular Oncology Laboratory, CIBERONC, Hospital Clinico San Carlos, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain. ⁵⁸SOD Genetica Molecolare, University Hospital, Pisa, Italy. ⁵⁹Department of Internal Medicine, Huntsman Cancer Institute, Salt Lake City, UT, USA. ⁶⁰Research Department, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia. ⁶¹Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Victoria, Australia. ⁶²Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁶³Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Cambridge, UK. ⁶⁴Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA. ⁶⁵Oncology

and Genetics Unit, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Xerencia de Xestión Integrada de Vigo-SERGAS, Vigo, Spain. ⁶⁶Axe Oncologie, Centre de Recherche, Centre Hospitalier Universitaire de Québec, Université Laval, Québec, Québec, Canada. ⁶⁷Departments of Pediatrics and Medicine, Columbia University, New York, NY, USA. ⁶⁸Centre for Medical Genetics, Ghent University, Ghent, Belgium. ⁶⁹Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia. ⁷⁰Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁷¹Division of Molecular Pathology, Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ⁷²Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ⁷³Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK. ⁷⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁷⁵Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA. ⁷⁶Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands. ⁷⁷Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. ⁷⁸Oncogenetics Group, Vall d'Hebron Institute of Oncology, Barcelona, Spain. ⁷⁹Clinical and Molecular Genetics Area, Vall Hebron University Hospital, Barcelona, Spain. ⁸⁰Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA. ⁸¹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia. ⁸²Basser Center for BRCA, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA. ⁸³Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. ⁸⁴Département de Médecine Moléculaire, Faculté de Médecine, Centre de Recherche, Centre Hospitalier Universitaire de Québec, Laval University, Québec City, Québec, Canada. ⁸⁵Cáncer Hereditario, Instituto de Biología y Genética Molecular (IBGM), Universidad de Valladolid Centro Superior de Investigaciones Científicas (UVA-CSIC), Valladolid, Spain. ⁸⁶Southampton Clinical Trials Unit, Faculty of Medicine, University of Southampton, Southampton, UK. ⁸⁷Cancer Sciences Academic Unit, Faculty of Medicine, University of Southampton, Southampton, UK. ⁸⁸School of Life Sciences, University of Westminster, London, UK. ⁸⁹Faculty of Medicine, University of Southampton, Southampton, UK. ⁹⁰Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. ⁹¹Genomic Medicine, Division of Evolution and Genomic Sciences, The University of Manchester, Manchester Academic Health Science Centre, Manchester Universities Foundation Trust, St. Mary's Hospital, Manchester, UK. ⁹²Genomic Medicine, North West Genomics Hub, Manchester Academic Health Science Centre, Manchester Universities Foundation Trust, St. Mary's Hospital, Manchester, UK. ⁹³David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, University of California, Los Angeles, Los Angeles, CA, USA. ⁹⁴Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. ⁹⁵Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK. ⁹⁶Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium. ⁹⁷Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. ⁹⁸Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic. ⁹⁹Program in Cancer Genetics, Departments of Human Genetics and Oncology, McGill University, Montréal, Québec, Canada. ¹⁰⁰The Suzanne Levy-Gertner Oncogenetics Unit, Chaim Sheba Medical Center, Ramat Gan, Israel. ¹⁰¹Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel. ¹⁰²School of Public Health, Curtin University, Perth, Western Australia, Australia. ¹⁰³Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain. ¹⁰⁴Moores Cancer Center, University of California, San Diego, La Jolla, CA, USA. ¹⁰⁵Schools of Medicine and Public Health, Division of Cancer Prevention and Control Research, Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, Los Angeles, CA, USA. ¹⁰⁶Cancer Risk and Prevention Clinic, Dana-Farber Cancer Institute, Boston, MA, USA. ¹⁰⁷Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain. ¹⁰⁸Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece. ¹⁰⁹Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia. ¹¹⁰Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, Victoria, Australia. ¹¹¹Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Kansas City, KS, USA. ¹¹²Department of Medicine, McGill University, Montréal, Québec, Canada. ¹¹³Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, Québec, Canada. ¹¹⁴Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA. ¹¹⁵L'Unità Operativa di Anatomia Patologica, Ospedale di Circolo, ASST Sette Laghi, Varese, Italy. ¹¹⁶Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. ¹¹⁷Department of Surgery, Oulu University Hospital, University of Oulu, Oulu, Finland. ¹¹⁸Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CHUM), Université de Montréal, Montréal, Québec, Canada. ¹¹⁹Cancer and Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Paris, France. ¹²⁰Center for Hereditary Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ¹²¹Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ¹²²Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. ¹²³Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. ¹²⁴Department of Oncology, Södersjukhuset, Stockholm, Sweden. ¹²⁵Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹²⁶Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland. ¹²⁷Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland. ¹²⁸Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland. ¹²⁹Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore. ¹³⁰Department of Surgery, National University Health System, Singapore, Singapore. ¹³¹Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, the Netherlands. ¹³²School of Population and Global Health, The University of Western Australia, Perth, Western Australia, Australia. ¹³³Family Cancer Clinic, The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ¹³⁴Division of Cancer Sciences, University of Manchester, Manchester, UK. ¹³⁵Center for Medical Genetics, NorthShore University HealthSystem, Evanston, IL, USA. ¹³⁶The University of Chicago Pritzker School of Medicine, Chicago, IL, USA. ¹³⁷N.N. Petrov Institute of Oncology, St. Petersburg, Russia. ¹³⁸Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA. ¹³⁹Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. ¹⁴⁰Research Centre for Genetic Engineering and Biotechnology 'Georgi D. Efremov', Macedonian Academy of Sciences and Arts, Skopje, Republic of North Macedonia. ¹⁴¹Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland. ¹⁴²Parkville Familial Cancer Centre, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia. ¹⁴³Hematology, Oncology and Transfusion Medicine Center, Department of Molecular and Regenerative Medicine, Vilnius University Hospital Santariskiu Clinics, Vilnius, Lithuania. ¹⁴⁴State Research Institute Centre for Innovative Medicine, Vilnius, Lithuania. ¹⁴⁵Department of Medicine, Division of Hematology/Oncology, UPMC Hillman Cancer Center, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ¹⁴⁶Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. ¹⁴⁷Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. ¹⁴⁸Department of Oncology, Tampere University Hospital, Tampere University and Tampere Cancer Center, Tampere, Finland. ¹⁴⁹Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁵⁰Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. ¹⁵¹Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea. ¹⁵²Cancer Research Institute, Seoul National University, Seoul, Korea. ¹⁵³Faculty of Medicine, University of Heidelberg, Heidelberg, Germany. ¹⁵⁴David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California, Los Angeles, Los Angeles, CA, USA. ¹⁵⁵Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ¹⁵⁶Surgery, School of Medicine, National University

of Ireland, Galway, Ireland. ¹⁵⁷Department of Genetics and Fundamental Medicine, Bashkir State Medical University, Ufa, Russia. ¹⁵⁸Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. ¹⁵⁹Familial Cancer Service, Weatmead Hospital, Sydney, New South Wales, Australia. ¹⁶⁰Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. ¹⁶¹Department of Internal Medicine, Evangelische Kliniken Bonn, Johanniter Krankenhaus, Bonn, Germany. ¹⁶²Molecular Diagnostics Laboratory, INRASTES, National Centre for Scientific Research 'Demokritos', Athens, Greece. ¹⁶³Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. ¹⁶⁴Department of Histopathology and Cytology, Clinical Hospital 'Acibadem Sistina', Skopje, Republic of North Macedonia. ¹⁶⁵Hong Kong Hereditary Breast Cancer Family Registry, Cancer Genetics Centre, Happy Valley, Hong Kong. ¹⁶⁶Department of Surgery, The University of Hong Kong, Pok Fu Lam, Hong Kong. ¹⁶⁷Department of Surgery, Hong Kong Sanatorium and Hospital, Happy Valley, Hong Kong. ¹⁶⁸VIB Center for Cancer Biology, Vlaams Instituut voor Biotechnologie, Leuven, Belgium. ¹⁶⁹Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium. ¹⁷⁰Institut Curie, Paris, France. ¹⁷¹Mines ParisTech, Paris, France. ¹⁷²Genetic Epidemiology of Cancer Team, INSERM U900, Paris, France. ¹⁷³Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. ¹⁷⁴Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden. ¹⁷⁵Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK. ¹⁷⁶Institute of Population Health, University of Manchester, Manchester, UK. ¹⁷⁷Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden. ¹⁷⁸Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA. ¹⁷⁹Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. ¹⁸⁰Division of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan. ¹⁸¹Cancer Epidemiology Group, University Cancer Center Hamburg (UCC), University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ¹⁸²Anatomical Pathology, The Alfred Hospital, Melbourne, Victoria, Australia. ¹⁸³Department of Tumour Biology, INSERM U830, Paris, France. ¹⁸⁴Department of Gynecology and Obstetrics, University of Munich, Munich, Germany. ¹⁸⁵NRG Oncology, Statistics and Data Management Center, Roswell Park Cancer Institute, Buffalo, NY, USA. ¹⁸⁶Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. ¹⁸⁷Laboratory Medicine Program, University Health Network, Toronto, Ontario, Canada. ¹⁸⁸Radiation Oncology, Hospital Meixoeiro-XXI de Vigo, Vigo, Spain. ¹⁸⁹Women's College Research Institute, University of Toronto, Toronto, Ontario, Canada. ¹⁹⁰Department of Biochemistry and Molecular Medicine, University of California, Davis, Davis, CA, USA. ¹⁹¹Center for Genomic Medicine at Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. ¹⁹²Latvian Biomedical Research and Study Centre, Riga, Latvia. ¹⁹³Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. ¹⁹⁴Clinical Genetics Research Laboratory, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹⁹⁵Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹⁹⁶Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary. ¹⁹⁷Center for Clinical Cancer Genetics, The University of Chicago, Chicago, IL, USA. ¹⁹⁸Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden. ¹⁹⁹Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, UK. ²⁰⁰University of New Mexico Health Sciences Center, University of New Mexico, Albuquerque, NM, USA. ²⁰¹Cancer Research UK Clinical Trials Unit, Institute of Cancer Sciences, University of Glasgow, Glasgow, UK. ²⁰²Molecular Diagnostics, Aalborg University Hospital, Aalborg, Denmark. ²⁰³Clinical Cancer Research Center, Aalborg University Hospital, Aalborg, Denmark. ²⁰⁴Department of Clinical Medicine, Aalborg University, Aalborg, Denmark. ²⁰⁵Genome Diagnostics Program, IFOM—the FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy. ²⁰⁶Translational Research Laboratory, Bellvitge Biomedical Research Institute (IDIBELL), Catalan Institute of Oncology, CIBERONC, Barcelona, Spain. ²⁰⁷Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland. ²⁰⁸Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland. ²⁰⁹Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy. ²¹⁰School of Women's and Children's Health, Faculty of Medicine, University of New South Wales, Sydney, New South Wales, Australia. ²¹¹The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. ²¹²Clinical Genetics, Karolinska Institutet, Stockholm, Sweden. ²¹³Clalit National Israeli Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel. ²¹⁴Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA. ²¹⁵Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain. ²¹⁶Department of Oncology, University Hospital of Larissa, Larissa, Greece. ²¹⁷Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA. ²¹⁸Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain. ²¹⁹Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago de Compostela, Spain. ²²⁰School of Medicine, University of Western Australia, Perth, Western Australia, Australia. ²²¹Research Oncology, Guy's Hospital, King's College London, London, UK. ²²²Faculty of Information Technology, Monash University, Melbourne, Victoria, Australia. ²²³National Center for Tumor Diseases, University Hospital and German Cancer Research Center, Heidelberg, Germany. ²²⁴Network Aging Research, University of Heidelberg, Heidelberg, Germany. ²²⁵Division of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, New South Wales, Australia. ²²⁶Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia. ²²⁷Hunter Medical Research Institute, John Hunter Hospital, Newcastle, New South Wales, Australia. ²²⁸Clinical Cancer Genetics Program, Division of Human Genetics, Department of Internal Medicine, The Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA. ²²⁹Department of Internal Medicine, Division of Medical Oncology, University of Kansas Medical Center, Westwood, KS, USA. ²³⁰Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ²³¹School of Public Health, China Medical University, Taichung, Taiwan. ²³²Department of Obstetrics and Gynecology and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria. ²³³Clinical Cancer Genomics, City of Hope, Duarte, CA, USA. ²³⁴University Clinic of Radiotherapy and Oncology, Medical Faculty, Ss. Cyril and Methodius University in Skopje, Skopje, Republic of North Macedonia. ²³⁵Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia. ²³⁶Population Oncology, BC Cancer, Vancouver, British Columbia, Canada. ²³⁷School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada. ²³⁸The Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and University of Western Australia, Perth, Western Australia, Australia. ²³⁹Service de Génétique, Institut Curie, Paris, France. ²⁴⁰Université Paris Descartes, Paris, France. ²⁴¹Institute of Human Genetics, University Hospital Heidelberg, Heidelberg, Germany. ²⁴²Division of Breast Cancer Research, The Institute of Cancer Research, London, UK. ²⁴³Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ²⁴⁴Department of Obstetrics and Gynecology, Medical University of Vienna, Vienna, Austria. ²⁴⁵Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA. ²⁴⁶Department of Genetics, Portuguese Oncology Institute, Porto, Portugal. ²⁴⁷Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal. ²⁴⁸Cancer Center, Kuopio University Hospital, Kuopio, Finland. ²⁴⁹Institute of Clinical Medicine, Oncology, University of Eastern Finland, Kuopio, Finland. ²⁵⁰Breast Cancer Research Programme, Cancer Research Malaysia, Kuala Lumpur, Malaysia. ²⁵¹Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. ²⁵²Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA. ²⁵³Hereditary Cancer Program, ONCOBELL-IDIBELL-IDIBGI-IGTP, Catalan Institute of Oncology, CIBERONC, Barcelona, Spain. ²⁵⁴Department of Clinical Genetics, Odense University Hospital, Odense, Denmark. ²⁵⁵Department of Medicine, Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ²⁵⁶Department of Medical Genetics, University of Cambridge, Cambridge, UK. ²⁵⁷Department of Cancer Biology and Genetics, The Ohio State University, Columbus, OH, USA. ²⁵⁸Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands. ²⁵⁹Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK. ²⁶⁰Wellcome Trust Centre for Human Genetics and NIHR Oxford Biomedical Research

Centre, University of Oxford, Oxford, UK. ²⁶¹Center for Population Health Research, National Institute of Public Health, Cuernavaca, Mexico. ²⁶²Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ²⁶³Department of Medical Oncology, Beth Israel Deaconess Medical Center, Boston, MA, USA. ²⁶⁴Department of Pathology, University Hospital of Heraklion, Heraklion, Greece. ²⁶⁵Frauenklinik der Stadtklinik Baden-Baden, Baden-Baden, Germany. ²⁶⁶Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA. ²⁶⁷Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. ²⁶⁸Department of Genetics, University of Pretoria, Pretoria, South Africa. ²⁶⁹Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica-USC, CIBERER, IDIS, Santiago de Compostela, Spain. ²⁷⁰Division of Functional Onco-genomics and Genetics, Centro di Riferimento Oncologico di Aviano (CRO), IRCCS, Aviano, Italy. ²⁷¹Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, USA. ²⁷²Department of Surgical Sciences, Uppsala University, Uppsala, Sweden. ²⁷³Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK. ²⁷⁴Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ²⁷⁵Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway. ²⁷⁶Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. ²⁷⁷The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON) Coordinating Center, The Netherlands Cancer Institute, Amsterdam, the Netherlands. ²⁷⁸Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia. ⁴⁵⁸These authors contributed equally: Hugues Aschard, Jonathan Beesley. ⁴⁵⁹These authors jointly supervised this work: Georgia Chenevix-Trench, Jacques Simard, Douglas F. Easton, Peter Kraft, Alison M. Dunning. ⁴⁶⁰A full list of members and affiliations appears at the end of the paper. *e-mail: pkraft@hsph.harvard.edu; amd24@medschl.cam.ac.uk

GEMO Study Collaborators

Véronique Mari²⁸², Pascaline Berthet²⁸³, Laurent Castera²⁸³, Dominique Vaur²⁸³, Hakima Lallaoui²⁸⁴, Yves-Jean Bignon²⁸⁵, Nancy Uhrhammer²⁸⁵, Valérie Bonadona²⁸⁶, Christine Lasset²⁸⁶, Françoise Révillion²⁸⁷, Paul Vennin²⁸⁷, Daniele Muller²⁸⁸, Denise Molina Gomes²⁸⁹, Olivier Ingster²⁹⁰, Isabelle Coupier²⁹¹, Pascal Pujol²⁹¹, Marie-Agnès Collonge-Rame²⁹², Isabelle Mortemousque²⁹³, Odile Bera²⁹⁴, Mickaëlle Rose²⁹⁴, Amandine Baurand²⁹⁵, Geoffrey Bertolone²⁹⁵, Laurence Faivre²⁹⁵, Hélène Dreyfus²⁹⁶, Dominique Leroux²⁹⁶, Laurence Venat-Bouvet²⁹⁷, Stéphane Bézieau²⁹⁸, Capucine Delnatte²⁹⁸, Jean Chiesa²⁹⁹, Brigitte Gilbert-Dussardier³⁰⁰, Paul Gesta³⁰¹, Fabienne Prieur Prieur³⁰², Myriam Bronner³⁰³, Johanna Sokolowska³⁰³, Florence Coulet³⁰⁴, Nadia Boutry-Kryza³⁰⁵, Alain Calender³⁰⁵, Sophie Giraud³⁰⁵, Mélanie Leone³⁰⁵, Sandra Fert-Ferrer³⁰⁶, Dominique Stoppa-Lyonnet³⁰⁷, Yue Jiao³⁰⁸, Fabienne Lesueur Lesueur³⁰⁸, Noura Mebirouk³⁰⁸, Emmanuelle Barouk-Simonet³⁰⁹, Virginie Bubien³⁰⁹, Michel Longy³⁰⁹, Nicolas Sevenet³⁰⁹, Laurence Gladieff³¹⁰, Christine Toulas³¹⁰, Audrey Reimineras³¹¹, Hagay Sobol³¹¹, Brigitte Bressac-de Pailleters³¹², Odile Cabaret³¹², Olivier Caron³¹², Marine Guillaud-Bataille³¹², Etienne Rouleau³¹², Muriel Belotti²⁴¹, Bruno Buecher²⁴¹, Sandrine Caputo²⁴¹, Chrystelle Colas³¹³, Antoine De Pauw²⁴¹, Emmanuelle Fourme³¹⁴, Marion Gauthier-Villars²⁴¹, Lisa Golmard²⁴¹, Virginie Moncoutier²⁴¹ and Claire Saule²⁴¹

²⁸²Centre Antoine Lacassagne, Nice, France. ²⁸³Centre François Baclesse, Caen, France. ²⁸⁴Centre Hospitalier de La Rochelle, La Rochelle, France. ²⁸⁵Centre Jean Perrin, Clermont-Ferrand, France. ²⁸⁶Centre Léon Bérard, Lyon, France. ²⁸⁷Centre Oscar Lambret, Lille, France. ²⁸⁸Centre Paul Strauss, Strasbourg, France. ²⁸⁹CHI Poissy, Poissy, France. ²⁹⁰CHU Angers, Angers, France. ²⁹¹CHU Arnaud-de-Villeneuve, Montpellier, France. ²⁹²CHU Besançon, Besançon, France. ²⁹³CHU Bretonneau, Tours and Centre Hospitalier de Bourges, Orléans, France. ²⁹⁴CHU de Martinique, Fort de France, France. ²⁹⁵CHU Dijon, Dijon, France. ²⁹⁶CHU Grenoble, Grenoble, France. ²⁹⁷CHU Limoges, Limoges, France. ²⁹⁸CHU Nantes, Nantes, France. ²⁹⁹CHU Nîmes Carémeau, Nîmes, France. ³⁰⁰CHU Poitiers, Poitiers, France. ³⁰¹CHU Poitiers, Centre Hospitalier d'Angoulême and Centre Hospitalier de Niort, Poitiers, France. ³⁰²CHU Saint Etienne, Saint Etienne, France. ³⁰³CHU Vandoeuvre-les-Nancy, Vandoeuvre-les-, Nancy, France. ³⁰⁴Groupe Hospitalier Pitié-Salpêtrière, Paris, France. ³⁰⁵Centre Léon Bérard, Hospices Civils de Lyon, Lyon, France. ³⁰⁶Hôtel Dieu Centre Hospitalier, Chambéry, France. ³⁰⁷Inserm, U830, Service de Génétique, Institut Curie, Université Paris Descartes, Paris, France. ³⁰⁸Inserm, U900, Institut Curie, PSL University, Mines ParisTech, Paris, France. ³⁰⁹Institut Bergonié, Bordeaux, France. ³¹⁰Institut Claudius Regaud, Toulouse, France. ³¹¹Institut Paoli Calmettes, Marseille, France. ³¹²Service de Génétique, Gustave Roussy, Villejuif, France. ³¹³Service de Génétique, Institut Curie, Paris, France. ³¹⁴Service de Génétique, Institut Curie, Paris, France.

EMBRACE Collaborators

Alan Donaldson³¹⁵, Alex Murray³¹⁶, Angela Brady³¹⁷, Carole Brewer³¹⁸, Caroline Pottinger³¹⁹, Clare Miller³²⁰, David Gallagher³²¹, Helen Gregory³²², Jackie Cook³²³, Jacqueline Eason³²⁴,

Julian Adlard³²⁵, Julian Barwell³²⁶, Kai-Ren Ong³²⁷, Katie Snape³²⁸, Lisa Walker³²⁹, Louise Izatt³³⁰, Lucy Side³³¹, Marc Tischkowitz³³², Mark T. Rogers³³³, Mary E. Porteous³³⁴, Munaza Ahmed³³⁵, Patrick J. Morrison³³⁶, Paul Brennan³³⁷, Ros Eeles³³⁸ and Rosemarie Davidson³³⁹

³¹⁵Clinical Genetics Department, St Michael's Hospital, Bristol, UK. ³¹⁶All Wales Medical Genetics Service, Singleton Hospital, Swansea, UK. ³¹⁷North West Thames Regional Genetics Service, Kennedy Galton Centre, The North West London Hospitals NHS Trust, Harrow, UK. ³¹⁸Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK. ³¹⁹All Wales Medical Genetics Service, Glan Clwyd Hospital, Rhyl, UK. ³²⁰Department of Clinical Genetics, Alder Hey Hospital, Liverpool, UK. ³²¹Mater Private Hospital, Dublin, UK. ³²²North of Scotland Regional Genetics Service, NHS Grampian and University of Aberdeen, Aberdeen, UK. ³²³Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Sheffield, UK. ³²⁴Nottingham Clinical Genetics Service, Nottingham University Hospitals NHS Trust, Nottingham, UK. ³²⁵Yorkshire Regional Genetics Service, Chapel Allerton Hospital, Leeds, UK. ³²⁶Leicestershire Clinical Genetics Service, University Hospitals of Leicester NHS Trust, Leicester, UK. ³²⁷West Midlands Regional Genetics Service, Birmingham Women's Hospital Healthcare NHS Trust, Birmingham, UK. ³²⁸Medical Genetics Unit, St George's, University of London, London, UK. ³²⁹Oxford Regional Genetics Service, Churchill Hospital, Oxford, UK. ³³⁰Clinical Genetics, Guy's and St. Thomas' NHS Foundation Trust, London, UK. ³³¹Wessex Clinical Genetics Service, The Princess Anne Hospital, Southampton, UK. ³³²Department of Medical Genetics, Addenbrooke's Treatment Centre, Addenbrooke's Hospital, Cambridge, UK. ³³³All Wales Medical Genetics Service, University Hospital of Wales, Cardiff, UK. ³³⁴South East of Scotland Regional Genetics Service, Western General Hospital, Edinburgh, UK. ³³⁵North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Trust, London, UK. ³³⁶Northern Ireland Regional Genetics Centre, Belfast City Hospital, Belfast, UK. ³³⁷Institute of General Medicine, International Centre for Life, Newcastle, UK. ³³⁸Oncogenetics Team, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Sutton, UK. ³³⁹Department of Clinical Genetics, South Glasgow University Hospitals, Glasgow, UK.

KConFab Investigators

Adrienne Sexton³⁴⁰, Alice Christian³⁴¹, Alison Trainer³⁴², Allan Spigelman³⁴³, Andrew Fellows³⁴⁴, Andrew Shelling³⁴⁵, Anna De Fazio³⁴⁶, Anneke Blackburn³⁴⁷, Ashley Crook³⁴⁸, Bettina Meiser³⁴⁹, Briony Patterson³⁵⁰, Christine Clarke³⁵¹, Christobel Saunders³⁵², Clare Hunt³⁵³, Clare Scott³⁵⁴, David Amor³⁵⁵, Deb Marsh³⁵⁶, Edward Edkins³⁵⁷, Elizabeth Salisbury³⁵⁸, Eric Haan³⁵⁹, Eveline Neidermayr³⁴², Finlay Macrea³⁶⁰, Gelareh Farshid³⁶¹, Geoff Lindeman³⁶², Georgia Trench³⁶³, Graham Mann³⁶⁴, Graham Giles³⁶⁵, Grantley Gill³⁶⁶, Heather Thorne³⁴², Ian Campbell³⁶⁷, Ian Hickie³⁶⁸, Ingrid Winship³⁶⁹, James Flanagan³⁷⁰, James Kollias³⁷¹, Jane Visvader³⁷², Jennifer Stone³⁷³, Jessica Taylor³⁷⁴, Jo Burke³⁷⁵, Jodi Saunus³⁷⁶, John Forbes³⁷⁷, John Hopper³⁷⁸, Jonathan Beesley³⁷⁹, Judy Kirk³⁸⁰, Juliet French³⁸¹, Kathy Tucker³⁸², Kathy Wu³⁸³, Kelly Phillips³⁸⁴, Lara Lipton³⁸⁵, Leslie Andrews³⁸⁶, Lizz Lobb³⁸⁷, Logan Walker³⁸⁸, Maira Kentwell³⁸⁹, Mandy Spurdle³⁹⁰, Margaret Cummings³⁹¹, Margaret Gleeson³⁹², Marion Harris³⁹³, Mark Jenkins³⁹⁴, Mary Anne Young³⁹⁵, Martin Delatycki³⁵⁵, Mathew Wallis³⁹⁶, Matthew Burgess³⁹⁷, Melanie Price³⁹⁸, Melissa Brown³⁹⁹, Melissa Southey⁴⁰⁰, Michael Bogwitz⁴⁰¹, Michael Field⁴⁰², Michael Friedlander⁴⁰³, Michael Gattas⁴⁰⁴, Mona Saleh⁴⁰⁵, Nick Hayward⁴⁰⁶, Nick Pachter⁴⁰⁷, Paul Cohen⁴⁰⁸, Pascal Duijf⁴⁰⁹, Paul James⁴¹⁰, Pete Simpson⁴¹¹, Peter Fong⁴¹², Phyllis Butow⁴¹³, Rachael Williams³⁴³, Rick Kefford⁴¹⁴, Rodney Scott⁴¹⁵, Roger Milne³⁶⁵, Rosemary Balleine⁴¹⁶, Sarah-Jane Dawson³⁴⁴, Sheau Lok⁴¹⁷, Shona O'Connell⁴¹⁸, Sian Greening⁴¹⁹, Sophie Nightingale⁴²⁰, Stacey Edwards³⁷⁹, Stephen Fox⁴²¹, Sue-Anne McLachlan⁴²², Sunil Lakhani⁴²³ and Yoland Antill⁴²⁴

³⁴⁰Familial Cancer Centre, Royal Melbourne Hospital, Melbourne, Victoria, Australia. ³⁴¹Genetics Department, Central Region Genetics Service, Wellington Hospital, Wellington, New Zealand. ³⁴²The Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. ³⁴³Family Cancer Clinic, St Vincent's Hospital, Sydney, New South Wales, Australia. ³⁴⁴The Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. ³⁴⁵Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand. ³⁴⁶Department of Gynaecological Oncology, Westmead Institute for Cancer Research, Westmead Hospital, Sydney, New South Wales, Australia. ³⁴⁷Australian National University, Canberra, Australian Capital Territory, Australia. ³⁴⁸Department of Clinical Genetics, Royal North Shore Hospital, Sydney, New South Wales, Australia. ³⁴⁹Prince of Wales Hospital, The University of New South Wales, Sydney, New South Wales, Australia. ³⁵⁰Clinical Genetics Service, Royal Hobart Hospital, Hobart Tasmania, Australia. ³⁵¹Westmead Institute for Cancer Research, University of Sydney, Westmead Hospital, Sydney, New South Wales, Australia. ³⁵²School of Surgery and Pathology, Queen Elizabeth II Medical Centre, Perth, Western Australia, Australia. ³⁵³Southern Health Familial Cancer Centre, Monash Medical Centre, Melbourne, Victoria, Australia. ³⁵⁴The Walter and Eliza Hall Institute of Medical Research, c/o Royal Melbourne Hospital, Melbourne, Victoria, Australia. ³⁵⁵Genetic Health Services Victoria, Royal Children's Hospital, Melbourne, Victoria, Australia. ³⁵⁶Kolling Institute of Medical Research, Royal North Shore Hospital, Sydney, New South Wales, Australia. ³⁵⁷Clinical Chemistry, Princess Margaret Hospital for Children, Perth, Western Australia, Australia. ³⁵⁸Anatomical Pathology, Prince of Wales Hospital, Sydney, New South Wales, Australia. ³⁵⁹Department of Medical Genetics, Women's and Children's Hospital, Adelaide, South Australia, Australia. ³⁶⁰Family Cancer Clinic, The Royal Melbourne

549 Hospital, Melbourne, Victoria, Australia. ³⁶¹SA Tissue Pathology, IMVS, Adelaide, South Australia, Australia. ³⁶²Breast Cancer Laboratory, The Walter and
 550 Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Melbourne, Victoria, Australia. ³⁶³Queensland Institute of Medical Research, Royal
 551 Brisbane Hospital, Brisbane, Queensland, Australia. ³⁶⁴Westmead Institute for Cancer Research, Westmead Millennium Institute, Sydney, New South
 552 Wales, Australia. ³⁶⁵Anti-Cancer Council of Victoria, Melbourne, Victoria, Australia. ³⁶⁶Department of Surgery, Royal Adelaide Hospital, Adelaide, South
 553 Australia, Australia. ³⁶⁷Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. ³⁶⁸Brain and Mind Centre, Sydney, New South Wales, Australia.
 554 ³⁶⁹Department of Medicine, Royal Melbourne Hospital, Melbourne, Victoria, Australia. ³⁷⁰Epigenetics Unit, Department of Surgery and Oncology, Imperial
 555 College, London, England. ³⁷¹Breast Endocrine and Surgical Unit, Royal Adelaide Hospital, Adelaide, South Australia, Australia. ³⁷²The Walter and Eliza Hall
 556 Institute of Medical Research, c/o Royal Melbourne Hospital, Melbourne, Victoria, Australia. ³⁷³Centre for Genetic Origins of Health and Disease,
 557 University of Western Australia, Perth, Western Australia, Australia. ³⁷⁴Familial Cancer and Genetics Medicine, Royal Melbourne Hospital, Melbourne,
 558 Victoria, Australia. ³⁷⁵Royal Hobart Hospital, Hobart, Tasmania, Australia. ³⁷⁶Breast Pathology, University of Queensland Centre for Clinical Research, Royal
 559 Brisbane and Women's Hospital, Brisbane, Queensland, Australia. ³⁷⁷Surgical Oncology, University of Newcastle, Newcastle Mater Hospital, Newcastle,
 560 New South Wales, Australia. ³⁷⁸Centre for MEGA Epidemiology, Melbourne, Victoria, Australia. ³⁷⁹Queensland Institute of Medical Research, Brisbane,
 561 Queensland, Australia. ³⁸⁰Familial Cancer Service, Department of Medicine, Westmead Hospital, Sydney, New South Wales, Australia. ³⁸¹School of
 562 Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland, Australia. ³⁸²Hereditary Cancer Clinic, Prince of Wales Hospital, Sydney,
 563 New South Wales, Australia. ³⁸³Family Cancer Clinic, St Vincent's Hospital, Sydney, Australia. ³⁸⁴Department of Medical Oncology, Peter MacCallum
 564 Cancer Centre, Melbourne, Victoria, Australia. ³⁸⁵Medical Oncology and Clinical Haematology Unit, Western Hospital, Melbourne, Victoria, Australia.
 565 ³⁸⁶Hereditary Cancer Clinic, Prince of Wales Hospital, Sydney, New South Wales, Australia. ³⁸⁷School of Medicine, University of Notre Dame, Sydney, New
 566 South Wales, Australia. ³⁸⁸Department of Pathology, University of Otago, Christchurch, New Zealand. ³⁸⁹The Family Cancer Clinic, Royal Melbourne
 567 Hospital, Melbourne, Victoria, Australia. ³⁹⁰Queensland Institute of Medical Research, Brisbane, Queensland, Australia. ³⁹¹Department of Pathology,
 568 University of Queensland Medical School, Brisbane, Queensland, Australia. ³⁹²Hunter Family Cancer Service, Sydney, New South Wales, Australia. ³⁹³Family
 569 Cancer Clinic, Monash Medical Centre, Melbourne, Victoria, Australia. ³⁹⁴Centre for MEGA Epidemiology, The University of Melbourne, Melbourne,
 570 Victoria, Australia. ³⁹⁵The Garvan Institute of Medical Research, The Kinghorn Cancer Centre, Sydney, New South Wales, Australia. ³⁹⁶The Family Cancer
 571 Clinic, Austin Health, Melbourne, Victoria, Australia. ³⁹⁷Clinical Genetics Service, Austin Health, Melbourne, Victoria, Australia. ³⁹⁸Medical Psychology,
 572 University of Sydney, Sydney, New South Wales, Australia. ³⁹⁹University of Queensland, Brisbane, Queensland, Australia. ⁴⁰⁰Genetic Epidemiology
 573 Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. ⁴⁰¹Familial Cancer Centre, The Royal Melbourne
 574 Hospital, Melbourne, Victoria, Australia. ⁴⁰²Clinical Genetics, Royal North Shore Hospital, Sydney, New South Wales, Australia. ⁴⁰³Department of Medical
 575 Oncology, Prince of Wales Hospital, Sydney, New South Wales, Australia. ⁴⁰⁴Queensland Clinical Genetic Service, Royal Children's Hospital, Brisbane,
 576 Queensland, Australia. ⁴⁰⁵Centre for Genetic Education, Prince of Wales Hospital, Sydney, New South Wales, Australia. ⁴⁰⁶Queensland Institute for Medical
 577 Research, Royal Brisbane Hospital, Brisbane, Queensland, Australia. ⁴⁰⁷Genetic Services of WA, King Edward Memorial Hospital, Perth, Western Australia,
 578 Australia. ⁴⁰⁸Gynaecological Cancer Research, St John of God Subiaco Hospital, Perth, Western Australia, Australia. ⁴⁰⁹The University of Queensland
 579 Diamantina Institute, Brisbane, Queensland, Australia. ⁴¹⁰Family Cancer Clinic, The Peter MacCallum Cancer Centre, Melbourne, Australia. ⁴¹¹The
 580 University of Queensland, Brisbane, Queensland, Australia. ⁴¹²Regional Cancer and Blood Services Auckland City Hospital, Auckland, New Zealand.
 581 ⁴¹³Medical Psychology Unit, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia. ⁴¹⁴Medical Oncology Department, Westmead Hospital,
 582 Sydney, New South Wales, Australia. ⁴¹⁵Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales, Australia. ⁴¹⁶Department of
 583 Medical Oncology, Westmead Hospital, Sydney, New South Wales, Australia. ⁴¹⁷Department of Medical Oncology, The Royal Melbourne Hospital,
 584 Melbourne, Victoria, Australia. ⁴¹⁸Southern Health Familial Cancer Centre, Melbourne, Victoria, Australia. ⁴¹⁹Illawarra Cancer Centre, Wollongong Hospital,
 585 Wollongong, New South Wales, Australia. ⁴²⁰Western Health and Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. ⁴²¹Pathology
 586 Department, Peter MacCallum Cancer Centre, VCCC Building, Melbourne, Victoria, Australia. ⁴²²Department of Oncology, St Vincent's Hospital,
 587 Melbourne, Victoria, Australia. ⁴²³UQ Centre for Clinical Research, University of Queensland, The Royal Brisbane and Women's Hospital Herston, Brisbane,
 588 Queensland, Australia. ⁴²⁴The Family Cancer Clinic, Cabrini Hospital, Melbourne, Victoria, Australia.

HEBON Investigators

589 **Cora Aalfs⁴²⁵, Hanne Meijers-Heijboer⁴²⁶, Klaartje van Engelen⁴²⁵, Hans Gille⁴²⁶, Ingrid Boere⁴²⁷,**
 590 **Margriet Collée⁴²⁸, Carolien van Deurzen⁴²⁹, Maartje Hooning⁴²⁷, Inge-Marie Obdeijn⁴³⁰,**
 591 **Ans van den Ouweland⁴²⁸, Caroline Seynaeve⁴²⁷, Sabine Siesling⁴³¹,**
 592 **Janneke Verloop⁴³¹, Christi van Asperen⁴³², Peter Devilee⁴³³, Twiggy van Cronenburg⁴³²,**
 593 **Rien Blok⁴³⁴, Maaïke de Boer⁴³⁴, Encarna Gómez García⁴³⁴, Muriel Adank⁴³⁵, Frans Hogervorst⁴³⁵,**
 594 **Denise Jenner⁴³⁶, Flora van Leeuwen⁴³⁶, Matti Rookus⁴³⁶, Nicola Russell⁴³⁷, Marjanka Schmidt⁴³⁸,**
 595 **Sandra van den Belt-Dusebout⁴³⁹, Carolien Kets⁴⁴⁰, Arjen Mensenkamp⁴⁴⁰, Truuske de Bock⁴⁴¹,**
 596 **Annemieke van der Hout⁴⁴², Marian Mourits⁴⁴³, Jan Oosterwijk⁴⁴², Margreet Ausems⁴⁴⁴ and**
 597 **Marco Koudijs⁴⁴⁴**

606 ⁴²⁵Department of Clinical Genetics, Academic Medical Center, Amsterdam, the Netherlands. ⁴²⁶Department of Clinical Genetics, VU University
 607 Medical Centre, Amsterdam, the Netherlands. ⁴²⁷Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, the
 608 Netherlands. ⁴²⁸Department of Clinical Genetics, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁴²⁹Department
 609 of Pathology, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁴³⁰Department of Radiology, Family Cancer
 610 Clinic, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁴³¹The Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, the
 611 Netherlands. ⁴³²Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. ⁴³³Department of Human Genetics and
 612 Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands. ⁴³⁴Department of Clinical Genetics, Maastricht University Medical
 613 Center, Maastricht, the Netherlands. ⁴³⁵Family Cancer Clinic, Netherlands Cancer Institute, Amsterdam, the Netherlands. ⁴³⁶Department of Epidemiology,
 614

615 Netherlands Cancer Institute, Amsterdam, the Netherlands. ⁴³⁷Department of Radiotherapy, Netherlands Cancer Institute, Amsterdam, the Netherlands.
 616 ⁴³⁸Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, the Netherlands. ⁴³⁹The Nationwide Network and
 617 Registry of Histo- and Cytopathology (PALGA), Houten, the Netherlands. ⁴⁴⁰Department of Human Genetics, Radboud University Medical Center,
 618 Nijmegen, the Netherlands. ⁴⁴¹Department of Oncological Epidemiology, University Medical Center, Groningen University, Groningen, The Netherlands.
 619 ⁴⁴²Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands. ⁴⁴³Department of Gynaecological Oncology, University
 620 Medical Center, Groningen University, Groningen, the Netherlands. ⁴⁴⁴Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the
 621 Netherlands.

622 623 **ABCTB Investigators**

624
625
626 **Christine Clarke⁴⁴⁵, Deborah Marsh⁴⁴⁶, Rodney Scott^{447,448}, Robert Baxter⁴⁴⁹, Desmond Yip^{450,451},**
 627 **Jane Carpenter⁴⁵², Alison Davis^{453,454}, Nirmala Pathmanathan^{455,456}, Peter Simpson⁴⁵⁷,**
 628 **Dinny Graham⁴⁴⁵ and Mythily Sachchithanathan⁴⁴⁵**
 629
630

631
632 ⁴⁴⁵Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, New South Wales, Australia. ⁴⁴⁶University
 633 of Technology Sydney, Translational Oncology Group, School of Life Sciences, Faculty of Science, Sydney, New South Wales, Australia. ⁴⁴⁷School of
 634 Biomedical Sciences, University of Newcastle, Newcastle, UK. ⁴⁴⁸Hunter Medical Research Institute and NSW Health Pathology North, Newcastle,
 635 Australia. ⁴⁴⁹Kolling Institute of Medical Research, University of Sydney, Sydney, New South Wales, Australia. ⁴⁵⁰Epigenetics and Transcription Laboratory,
 636 Melanie Swan Memorial Translational Centre, Sci-Tech, University of Canberra, Canberra, Australian Capital Territory, Australia. ⁴⁵¹Department of
 637 Medical Oncology, The Canberra Hospital, Canberra, Australian Capital Territory, Australia. ⁴⁵²Scientific Platforms, The Westmead Institute for Medical
 638 Research, The University of Sydney, Sydney, New South Wales, Australia. ⁴⁵³The Canberra Hospital, Canberra, Australian Capital Territory, Australia.
 639 ⁴⁵⁴The Australian National University, Canberra, Australian Capital Territory, Australia. ⁴⁵⁵Westmead Breast Cancer Institute, Western Sydney Local Health
 640 District, Sydney, New South Wales, Australia. ⁴⁵⁶University of Sydney, Western Clinical School, Sydney, New South Wales, Australia. ⁴⁵⁷UQ Centre for
 641 Clinical Research, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia.
 642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680

Methods

Study samples. Epidemiological data for European women were obtained from 75 breast cancer case-control studies participating in the BCAC (cases: 40,285 iCOGS and 69,615 OncoArray; cases with ER status available: 29,561 iCOGS and 55,081 OncoArray; controls: 38,058 iCOGS and 50,879 OncoArray). Details of the participating studies, genotype calling and quality control are given in refs. ^{2,22,23}, respectively. Epidemiological data for *BRCA1* mutation carriers were obtained from 60 studies providing data to the CIMBA (affected: 1,591 iCOGS and 7,772 OncoArray; unaffected: 1,665 iCOGS and 7,780 OncoArray). This dataset has been described in detail previously^{1,59,60}. All studies provided samples of European ancestry. Any non-European samples were excluded from the analyses.

Variant selection and genotyping. Similar approaches were used to select variants for inclusion on the iCOGS and OncoArray, and these are described in detail elsewhere^{2,21}. Both arrays included a dense coverage of variants across known susceptibility regions (at the time of their design), with sparser coverage of the rest of the genome. Twenty-one known susceptibility regions were selected for dense genotyping using iCOGS and 73 regions were selected for OncoArray. These regions were 1-megabase (Mb) intervals centered on the published lead GWAS hit (combined into larger intervals where these overlapped). For iCOGS, all known variants from the March 2010 release of the 1000 Genomes Project with a MAF > 0.02 in Europeans were identified, and all those correlated with the published GWAS variants at $r^2 > 0.1$, together with a set of variants designed to tag all remaining variants at $r^2 > 0.9$, were selected to be included in the array (http://ccge.medschl.cam.ac.uk/files/2014/03/iCOGS_detailed_lists_ALL1.pdf). For OncoArray, all designable variants correlated with the known hits at $r^2 > 0.6$, plus all variants from lists of potentially functional variants on RegulomeDB and a set of variants designed to tag all of the remaining variants at $r^2 > 0.9$, were selected. In total, across the 152 regions considered here, 26,978 iCOGS- and 58,339 OncoArray-genotyped variants passed the quality control criteria.

We imputed genotypes for all of the remaining variants using IMPUTE2 (ref. ⁶¹) and the October 2014 release of the 1000 Genomes Project as a reference. Imputation was conducted independently in the iCOGS and OncoArray subsets. To improve accuracy at low-frequency variants, we used the standard IMPUTE2 MCMC algorithm for follow-up imputation, which includes no pre-phasing of the genotypes and increasing both the buffer regions and the number of haplotypes to use as templates (a more detailed description of the parameters used can be found in ref. ²¹). We thus genotyped or successfully imputed 639,118 variants (all with an imputation info score ≥ 0.3 and a MAF ≥ 0.001 in both iCOGS and OncoArray datasets). Imputation summaries and coverage for each of the analyzed regions stratified by allele frequency can be found in Supplementary Table 1b.

BCAC statistical analyses. Per-allele odds ratios and standard errors were estimated for each variant using logistic regression. We ran this analysis separately for iCOGS and OncoArray, and for overall, ER-positive and ER-negative breast cancer. The association between each variant and breast cancer risk was adjusted by study (iCOGS) or country (OncoArray), and eight (iCOGS) or ten (OncoArray) ancestry-informative principal components. The statistical significance for each variant was derived using a Wald test.

Defining appropriate significance thresholds for association signals. To establish an appropriate significance threshold for independent signals, all variants evaluated in the meta-analysis were included in logistic forward selection regression analyses for overall breast cancer risk in iCOGS, run independently for each region. We evaluated five *P* value thresholds for inclusion: $<1 \times 10^{-3}$, $<1 \times 10^{-5}$, $<1 \times 10^{-6}$, $<1 \times 10^{-7}$ and $<1 \times 10^{-8}$. The most parsimonious iCOGS models were tested in OncoArray, and the FDR at the 1% level for each threshold was estimated using the Benjamini-Hochberg procedure. At a 1% FDR threshold, 72% of associations, significant at $P < 10^{-4}$, were replicated on iCOGS, and 94% of associations, significant at $P < 10^{-6}$, were replicated on OncoArray. Based on these results, two categories were defined: strong-evidence signals (conditional $P < 10^{-6}$ in the final model) and moderate-evidence signals (conditional $P < 10^{-4}$ and $P \geq 10^{-6}$ in the final model).

Identification of independent signals. To identify independent signals, we ran multinomial stepwise regression analyses, separately in iCOGS and OncoArray, for all variants displaying evidence of association ($n_{\text{variants}} = 202,749$). We selected two sets of well-imputed variants (imputation info score ≥ 0.3 in both iCOGS and OncoArray): (1) common and low-frequency variants (MAF ≥ 0.01) with a logistic regression *P* value inclusion threshold of ≤ 0.05 in either the iCOGS or OncoArray datasets for at least one of the three phenotypes (overall, ER positive and ER negative breast cancer); and (2) rarer variants (MAF ≥ 0.001 and < 0.01), with a logistic regression inclusion *P* value of ≤ 0.0001 . The same parameters used for adjustment in logistic regression were used in the multinomial regression analysis (R function multinom). The multinomial regression estimates were combined using a fixed-effects meta-analysis weighted by the inverse variance. Variants with the lowest conditional *P* value from the meta-analysis of both European cohorts at each step were included in the multinomial regression model. However, if the new variant to be included in the model caused collinearity problems due to

high correlation with an already selected variant, or showed high heterogeneity ($P < 10^{-4}$) between iCOGS and OncoArray after being conditioned by the variant(s) in the model, we dropped the new variant and repeated this process.

At 105 of 152 evaluated regions, the main signal showed genome-wide significance, while 44 were marginally significant ($9.89 \times 10^{-5} \geq P > 5 \times 10^{-8}$). For two regions, there were no variants significant at $P < 10^{-4}$ (chr14:104712261–105712261; rs10623258; multinomial regression $P = 2.32 \times 10^{-4}$; chr19:10923703–11923703; rs322144; multinomial regression $P = 3.90 \times 10^{-3}$). Four main differences in the datasets used here and in the previous paper may account for this: (1) our previous paper² included data from 11 additional GWASs (14,910 cases and 17,588 controls) that have not been included in the present analysis in order to minimize differences in array coverage, and because ER status data were substantially incomplete and individual-level data were not available for all GWASs; (2) the present analysis was based on estimating separate risks for ER-positive and ER-negative disease, whereas in our previous paper the outcome was overall breast cancer risk; ER status was available for only 73% of the iCOGS and 79% of the OncoArray breast cancer cases; (3) for the set of samples genotyped with both arrays, ref. ² used the iCOGS genotypes, while the present study included OncoArray genotypes to maximize the number of samples genotyped with a larger coverage; and (4) the imputation procedure was modified (in particular using one-step imputation without pre-phasing) to improve the imputation accuracy of less frequent variants.

We used a forward stepwise approach to define the number of independent signals within each associated genomic region. First, we identified the index variant of the main signal in the region, and then ran multinomial logistic regression for all of the other variants, adjusted by the index variant, to identify additional variants that remained independently significant within the model. We repeated this process, adjusting for identified index variants, until no more additional variants could be added. In this way, we found from 1–11 independent signals within the 150 regions that containing a genome-wide significant main signal.

Selection of a set of CCVs. For each independently associated signal, we first defined CCVs likely to drive its association as those variants with *P* values within two orders of magnitude of the most significant variant for that signal, after adjusting for the index variant of other signals within that region (as identified in the forward stepwise regression above; Supplementary Fig. 6a)²⁴. For each region, we then attempted to obtain the best-fitting model by successively fitting models in which the index variant for each signal was replaced by other CCVs for that signal, adjusting for the index variants for the other signals (Supplementary Fig. 6b). Where a model with a higher chi-squared value was obtained, the index variant was replaced by the CCV in the best model (Supplementary Fig. 6c,d). This process was repeated until the model (that is, the set of index variants) did not change further (Supplementary Fig. 6g). This procedure was performed first for the set of strong signals (that is, considering models including only the strong signals). Once a final model had been obtained for the strong signals, the index variants for the strong signals were considered fixed and the process was repeated for all signals, the index variants for the weak signals (but not the strong signals) to vary. Using this procedure, we could define the best model for 140 out of 150 regions, but for ten regions this approach did not converge (chr4:175328036–176346426, chr5:55531884–56587883, chr6:151418856–152937016, chr8:75730301–76917937, chr10:80341148–81387721, chr10:122593901–123849324, chr12:115336522–116336522, chr14:36632769–37635752, chr16:3606788–4606788 and chr22:38068833–39859355). For these ten regions, we defined the best model, from among all possible combinations of credible variants, as that with the largest chi-squared value. Finally, we redefined the set of CCVs for each signal using the conditional *P* values, after adjusting for the revised set of index variants. Again, for the strong signals, we conditioned on the index variants for the other strong signals, while for the weak signals we conditioned on the index variants for all of the other signals.

Case-only analysis. Differences in the effect size between ER-positive and ER-negative disease for each index-independent variant were assessed using a case-only analysis. We performed logistic regression with ER status as the dependent variable and the lead variant at each strong signal in the fine-mapping region as the independent variables. We used FDR (5%) to adjust for multiple testing.

OncoArray-only stepwise analysis. To evaluate whether the lower coverage in iCOGS could affect the identification of independent signals, we ran stepwise multinomial regression using only the OncoArray dataset. We identified 249 independent signals. Ninety-two signals, in 67 fine-mapping regions, achieved a genome-wide significance level (conditional $P < 5 \times 10^{-8}$). Of these, 205 signals were also identified in the meta-analysis with iCOGS. Nine independent variants across ten regions were not evaluated in the combined analysis due to their low imputation information score in iCOGS. Of these nine signals, two signals would be classified as main primary signals: rs114709821 at region chr1:145144984–146144984 (OncoArray imputation information score = 0.72); and rs540848673 at region chr1:149406413–150420734 (OncoArray imputation information score = 0.33). Given the low number of additional signals identified in the OncoArray dataset alone, all analyses were based on the combined iCOGS/OncoArray dataset.

CIMBA statistical analysis. CIMBA provided data from 60 retrospective cohort studies consisting of 9,445 unaffected and 9,363 affected female *BRCA1* mutation carriers of European ancestry. Unconditional (that is, single-variant) analyses were performed using a score test based on the retrospective likelihood of observing the genotype conditional on the disease phenotype^{62,63}. Conditional analyses, where more than one variant is analyzed simultaneously, cannot be performed in this score test framework. Therefore, conditional analyses were performed by Cox regression, allowing for adjustment of the conditionally independent variants identified by the BCAC/DRIVE analyses. All models were stratified by country and birth cohort, and adjusted for relatedness (unconditional models used kinship-adjusted standard errors based on the estimated kinship matrix; conditional models used cluster robust standard errors based on phenotypic family data).

Data from the iCOGS array and OncoArray were analyzed separately and combined to give an overall *BRCA1* association by fixed-effects meta-analysis. Variants were excluded from further analyses if they exhibited evidence of heterogeneity (heterogeneity $P < 1 \times 10^{-4}$) between iCOGS and OncoArray, had a MAF < 0.005 , were poorly imputed (imputation information score < 0.3) or were imputed to iCOGS only (that is, they must have been imputed to OncoArray or iCOGS and OncoArray).

Meta-analysis of ER-negative cases in BCAC with *BRCA1* mutation carriers from CIMBA. *BRCA1* mutation carrier association results were combined with the BCAC multinomial regression ER-negative association results in a fixed-effects meta-analysis. Variants considered for analysis must have passed all previous quality control steps and have had MAF ≥ 0.005 . All meta-analyses were performed using the METAL software⁶⁴. Instances where spurious associations might occur were investigated by assessing the linkage disequilibrium between a possible spurious association and the conditionally independent variants. High linkage disequilibrium between a variant and a conditionally independent variant within its region causes model instability through collinearity, and convergence of the model likelihood maximization may not be reliable. Where the association appeared to be driven by collinearity, the signals were excluded.

Heritability estimation. To estimate the frailty-scale heritability due to all of the fine-mapping signals, we used the formula:

$$h^2 = 2(\gamma^T R \gamma - \tau^T I \tau)$$

Here, $\gamma = \gamma \sqrt{\mathbf{p}(1-\mathbf{p})}$ and $\tau^T = \tau \sqrt{\mathbf{p}(1-\mathbf{p})}$, where \mathbf{p} is a vector of allele frequencies, γ are the estimated per-allele odds ratios, τ are the corresponding standard errors and R is the correlation matrix of genotype frequencies.

To adjust for the overestimation resulting from only including signals passing a given significance threshold, we adapted the approach of ref. ⁶⁵, based on maximizing the likelihood conditional on the test statistic passing the relevant threshold. Since our analyses were based on estimating ER-negative and ER-positive odds ratios simultaneously, the method needed to be adapted to maximize a conditional bivariate normal likelihood. Following ref. ⁶⁵, we then estimated mean square error estimates based on a weighted mean of the maximum likelihood estimates and the naïve estimates, which were shown to be unbiased in the 1-degree of freedom case. The estimated effect sizes for overall breast cancer were computed as a weighted mean of the ER-negative and ER-positive estimates, based on the proportions of each subtype in the whole study (weights: 0.21 and 0.79). The results were then expressed in terms of the proportion of the FRR to first-degree relatives of affected women, using the formula $h^2/(2\log(\lambda))$, where the FRR λ was assumed to be 2 (ref. ²).

eQTL analysis. Total RNA was extracted from normal breast tissue in formalin-fixed paraffin-embedded breast cancer tissue blocks from 264 NHS participants³². Transcript expression levels were measured using the Glue Grant Human Transcriptome Array version 3.0 at the Molecular Biology Core Facilities, Dana-Farber Cancer Institute. Gene expression was normalized and summarized into log₂ values using RMA (Affymetrix Power Tools version 1.18.012). Quality control was performed using GlueQC and arrayQualityMetrics version 3.24.014. Genome-wide data on variants were generated using the Illumina HumanHap550 BeadChip as part of the Cancer Genetic Markers of Susceptibility initiative⁶⁶. Imputation to the 1000KGP Phase 3 version 5 ALL reference panel was performed using MACH to pre-phase measured genotypes, and minimac to impute.

Expression analyses were performed using data from the TCGA and METABRIC projects^{34,38}. The TCGA eQTL analysis was based on 458 breast tumors that had matched gene expression, copy number and methylation profiles, together with the corresponding germline genotypes available. All 458 individuals were of European ancestry, as ascertained using the genotype data and the Local Ancestry in Admixed Populations (LAMP) software package (LAMP estimate cut-off $> 95\%$ European)⁶⁷. Germline genotypes were imputed into the 1000 Genomes Project reference panel (October 2014 release) using IMPUTE version 2 (refs. ^{68,69}). Gene expression had been measured on the Illumina HiSeq 2000 RNA sequencing (RNA-Seq) platform (gene-level RSEM normalized counts⁷⁰), copy number estimates were derived from Affymetrix SNP 6.0 (somatic copy number alteration minus germline copy number variation called using the GISTIC2

algorithm⁷¹), and methylation beta values were measured on the Illumina Infinium HumanMethylation450. Expression QTL analysis focused on all variants within each of the 152 genomic intervals that had been subjected to fine-mapping for their association with breast cancer susceptibility. Each of these variants was evaluated for its association with the expression of every gene within 2 Mb that had been profiled for each of the three data types. The effects of tumor copy number and methylation on gene expression were first regressed out using a method described previously⁷². eQTL analysis was performed by linear regression, with residual gene expression as the outcome, germline SNP genotype dosage as the covariate of interest, and ESR1 expression and age as additional covariates, using the R package Matrix eQTL⁷³.

The METABRIC eQTL analysis was based on 138 normal breast tissue samples resected from patients with breast cancer of European ancestry. Germline genotyping for the METABRIC study was also done on the Affymetrix SNP 6.0 array, and gene expression in the METABRIC study was measured using the Illumina HT12 microarray platform (probe-level estimates). No adjustment was implemented for somatic copy number and methylation status since we were evaluating eQTLs in normal breast tissue. All other steps were identical to the TCGA eQTL analysis described above.

Genomic features enrichment. We explored the overlap of CCVs and excluded variants with 90 transcription factors, ten histone marks and DNase hypersensitivity sites in 15 breast cell lines and eight normal human breast tissues. We analyzed data from the Encyclopedia of DNA Elements (ENCODE) Project^{74,75}, Roadmap Epigenomics Projects⁷⁶, the International Human Epigenome Consortium^{77,77}, Pellacani et al.⁷⁸, TCGA³³, METABRIC³⁴, the ReMap database (we included 241 transcription factor annotations from ReMap (from a total of 2,825), which showed at least 2% overlap for any of the phenotype SNP sets)⁷⁹ and other data obtained through the National Center for Biotechnology Information Gene Expression Omnibus. Promoters were defined following the procedure defined in ref. ⁷⁸ (that is, ± 2 kilobases (kb) from a gene transcription start site) using an updated version of the RefSeq genes (refGene version updated 11 April 2017)⁸⁰. Transcribed regions were defined using the same version of RefSeq genes. lncRNA annotation was obtained from GENCODE (version 19)⁸¹.

To include eQTL results in the enrichment analysis we: (1) identified all of the genes for which summary statistics were available; (2) defined the most significant eQTL variant for each gene (index eQTL variant; P value threshold $\leq 5 \times 10^{-4}$); and (3) classified variants with P values within two orders of magnitude of the index expression variant as the credible set of eQTL variants (that is, the best candidates to drive expression of the gene). Variants within at least one eQTL credible set were defined as expression variants. We evaluated the overlap between eQTL credible sets and CCVs (risk variants credible set). We evaluated the enrichment of CCVs for genomic features using logistic regression, with CCV (versus non-CCV variants) being the outcome. To adjust for the correlation among variants in the same fine-mapping region, we used robust variance estimation for clustered observations (R function multiwaycov). The associated variants at an FDR of 5% were included in a stepwise forward logistic regression procedure to select the most parsimonious model. A likelihood ratio test was used to compare multinomial logistic regression models with and without equality effect constraints to evaluate whether there was heterogeneity among the effect sizes for ER positive, ER negative or signals equally associated with both phenotypes (ER neutral).

To validate the disease specificity of the regulatory regions identified through this analysis, we followed the same approach for the autoimmune-related CCVs from ref. ²⁹ ($n = 4,192$). Variants excluded as candidate causal variants, and within 500 kb upstream and downstream of the index variant for each signal, were classified as excluded variants ($n = 1,686,484$). We then tested the enrichment for both the breast cancer and autoimmune CCVs with breast and T and B cell enhancers. We also evaluated the overlap of our CCVs with ENCODE enhancer-like and promoter-like regions for 111 tissues, primary cells, immortalized cell lines and in vitro-differentiated cells. Of these, 73 had available data for both enhancer- and promoter-like regions.

Transcription binding site motif analysis. We conducted a search to find motif occurrences for the transcription factors significantly enriched in the genomic featured. For this, we used two publicly available databases: Factorbook⁸² and JASPAR 2016 (ref. ⁸³). For the search using Factorbook, we included the motifs for the transcription factors discovered in the cell lines where significant enrichment was found in our genomic features analysis. We also searched for all of the available motifs for *Homo sapiens* in the JASPAR database (JASPAR CORE 2016; TFBSTools⁸⁴). Using the USCS sequence (BSgenome.Hsapiens.USCS.hg19) as a reference, we created fasta sequences with the reference and alternative alleles for all of the variants included in our analysis plus 20 base pairs flanking each variant. We used FIMO (version 4.11.2; Grant et al.⁸⁵) to scan all of the fasta sequences, searching for the JASPAR and Factorbook motifs to identify any overlap of any of the alleles for each of the variants (setting the P value threshold to 10^{-3}). We subsequently determined whether our CCVs were more frequency overlapping a particular transcription factor binding motif when compared with the excluded variants. We ran these analyses for all of the strong signals, but also strong signals stratified by ER status. Also, we subset this analysis to the variants located at

813 regulatory regions in an ER-positive cell line (MCF-7 marked by H3K4me1;
814 ENCODE identification: ENCF674BKS) and evaluated whether the ER-positive
815 CCVs overlapped any of the motifs more frequently than the excluded variants. We
816 also evaluated the change in total binding affinity caused by the ER-positive CCCR
817 alternative allele for all but one (2:217955891:T:<CNO>:0) of the ER-positive CCVs
(MatrixRider⁶⁶).

818 Subsequently, we evaluated whether the MCF-7 regions demarked by H3K4me1
819 (ENCODE identification: ENCF674BKS) and overlapped by ER-positive CCVs
820 were enriched in known TFBS motifs. First, we subset the ENCODE bed file
821 ENCF674BKS to identify MCF-7 H3K4me1 peaks overlapped by the ER-positive
822 CCVs ($n = 107$), as well as peaks only overlapped by excluded variants ($n = 11,099$),
823 using BEDTools⁸⁷. We created fasta format sequences using genomic coordinate
824 data from the intersected bed files. To create a control sequence set, we used the
825 script included with the MEME Suite (fasta-shuffle-letters) to create ten shuffled
826 copies of each sequence overlapped by ER-positive CCVs ($n = 1,070$). We then
827 used AME⁸⁸ to interrogate whether the 107 MCF-7 H3K4me1 genomic regions
828 overlapped by ER-positive CCVs were enriched in known TFBS consensus motifs
829 when compared with the shuffled control sequences, or with the MCF-7 H3K4me1
830 genomic regions overlapped only by excluded variants. We used the command line
831 version of AME (version 4.12.0), selecting as a scoring method the total number of
832 positions in the sequence whose motif score P value was $<10^{-3}$, and using a one-
833 tailed Fisher's exact test as the association test.

832 **PAINTOR analysis.** To further refine the set of CCVs, we performed empirical
833 Bayes fine-mapping using PAINTOR to integrate marginal genetic association
834 summary statistics, linkage disequilibrium patterns and biological features^{31,89}.
835 PAINTOR derives jointly the posterior probability for causality of all variants
836 along the respective contribution of genomic features, in order to maximize the
837 log-likelihood of the data across all regions. PAINTOR does not assume a fixed
838 number of causal variants in each region, although it implicitly penalizes non-
839 parsimonious causal models. We applied PAINTOR separately to association
840 results for overall breast cancer (in 85 regions determined to have at least one
841 ER-neutral association or ER-positive and ER-negative association), ER-positive
842 breast cancer (in 48 regions determined to have at least one ER-positive-
843 specific association) and ER-negative breast cancer (in 22 regions determined
844 to have at least one ER-negative-specific association). To avoid artefacts due
845 to mismatches between the linkage disequilibrium in study samples and the
846 linkage disequilibrium matrix supplied to PAINTOR, we used association logistic
847 regression summary statistics from OncoArray data only, and estimated the linkage
848 disequilibrium structure in the OncoArray sample. For each endpoint, we fit four
849 models with increasing numbers of genomic features selected from the stepwise
850 enrichment analyses described above: model 0 (with no genomic features; assumes
851 each variant is equally likely to be causal a priori); model 1 (with those genomic
852 features selected with the stopping rule $P < 0.001$); model 2 (with those genomic
853 features selected with the stopping rule $P < 0.01$); and model 3 (with those genomic
854 features selected with the stopping rule $P < 0.05$).

852 We used the Bayesian information criterion (BIC) to choose the best-fitting
853 model for each outcome. As PAINTOR estimates the marginal log-likelihood
854 of the observed Z scores using Gibbs sampling, we used a shrunk mean BIC
855 across multiple Gibbs chains to account for the stochasticity in the log-likelihood
856 estimates. We ran PAINTOR four times to generate four independent Gibbs
857 chains, and estimated the BIC difference between model i and model j as
858 $\Delta_{ij} = \left(\frac{100}{i+j-100}\right) (BIC_i - BIC_j)$. This assumes a $n(0,100)$ prior on the difference, or
859 roughly a 16% chance that model i would be decisively better than model j (that
860 is $|BIC_i - BIC_j| > 10$). We then proceeded to choose the best-fitting model in a
861 stepwise fashion: starting with a model with no annotations, we selected a model
862 with more annotations in favor of a model with fewer if the larger model was a
863 considerably better fit (that is, $\Delta_{ij} > 2$). Model 1 was the best fit according to this
864 process for overall and ER-positive breast cancer, while model 0 was the best fit for
865 ER-negative breast cancer.

863 Differences between the PAINTOR and CCV outputs may be due to several
864 factors. By considering functional enrichment and joint linkage disequilibrium
865 among all SNPs, PAINTOR may refine the set of likely causal variants; rather than
866 imposing a hard threshold, PAINTOR allows for a gradient of evidence supporting
867 causality, and the two sets of calculations are based on different summary statistics.
868 CCV analyses used both iCOGS and OncoArray genotypes, while PAINTOR used
869 only OncoArray data (Fig. 1 and Methods).

869 **Variant annotation.** Variant genome coordinates were converted to assembly
870 GRCh38 with liftOver and uploaded to Variant Effect Predictor⁹⁰ to determine
871 their effect on genes, transcripts and protein sequence. The commercial software
872 Alamut Batch version 1.6 was also used to annotate coding and splicing variants.
873 PolyPhen-2 (ref. ⁹¹), SIFT⁹² and MAPP⁹³ were used to predict the consequences of
874 missense coding variants. MaxEntScan⁹⁴, Splice-Site Finder and Human Splicing
875 Finder⁹⁵ were used to predict splicing effects.

876 **INQUISIT analysis.** Logic underlying INQUISIT predictions. Briefly, genes were
877 considered to be potential targets of candidate causal variants through effects on:
878 (1) distal gene regulation; (2) proximal regulation; or (3) a gene's coding sequence.

828 We intersected CCV positions with multiple sources of genomic information,
829 including chromatin interactions from capture Hi-C experiments performed in
830 a panel of six breast cell lines⁹⁶, ChIA-PET⁹⁷ and Hi-C (Rao et al. 2014). We used
831 computational enhancer-promoter correlations (PreSTIGE⁹⁸, IM-PET (He et al.
832 2014), FANTOMS (ref. ⁹⁹) and super-enhancers⁹⁸), results for breast tissue-specific
833 expression variants from multiple independent studies (TCGA, METABRIC and
834 NHS; Methods), allele-specific imbalance in gene expression¹⁰⁰, transcription factor
835 and histone modification ChIP-Seq from the ENCODE and Roadmap Epigenomics
836 Projects, together with the genomic features found to be significantly enriched as
837 described above, gene expression RNA-Seq from several breast cancer lines and
838 normal samples, and topologically associated domain boundaries from T-47D
839 cells (ENCODE¹⁰¹; Methods and Key Resources Table). To assess the impact of
840 intragenic variants, we evaluated their potential to alter splicing using Alamut
841 Batch to identify new and cryptic donors and acceptors, and several tools to predict
842 the effects of coding sequence changes (see 'Variant annotation' section). Variants
843 potentially affecting post-translational modifications were downloaded from the 'A
844 Website Exhibits SNP On Modification Event' database (<http://www.awesome-hust.com/>)¹⁰². The output from each tool was converted to a binary measure to indicate
845 deleterious or tolerated predictions.

846 **Scoring hierarchy.** Each target gene prediction category (distal, promoter or coding)
847 was scored according to different criteria. Genes predicted to be distally regulated
848 targets of CCVs were awarded points based on physical links (for example, CHi-C),
849 computational prediction methods, allele-specific expression or expression variant
850 associations. All CCVs and HPPVs were considered as potentially involved in
851 distal regulation. Intersection of a putative distal enhancer with genomic features
852 found to be significantly enriched (see 'Genomic features enrichment' for details)
853 were further upweighted. Multiple independent interactions were awarded an
854 additional point. CCVs and HPPVs in gene proximal regulatory regions were
855 intersected with histone ChIP-Seq peaks characteristic of promoters and assigned to
856 the overlapping transcription start sites (defined as -1.0 kb to $+0.1$ kb). Further
857 points were awarded to such genes if there was evidence of expression variant
858 association or allele-specific expression, while a lack of expression resulted in
859 down-weighting as potential targets. Potential coding changes, including missense,
860 nonsense and predicted splicing alterations, resulted in the addition of one point
861 to the encoded gene for each type of change, while lack of expression reduced the
862 score. We added an additional point for predicted target genes that were also breast
863 cancer drivers. For each category, scores ranged from 0–7 (distal), 0–3 (promoter)
864 or 0–2 (coding). We converted these scores into 'confidence levels': level 1 (highest
865 confidence; distal score > 4 , promoter score ≥ 3 and coding score > 1); level 2
866 ($1 \leq$ distal score ≤ 4 , promoter score = 1 or 2 and coding score = 1); and level 3
867 ($0 <$ distal score < 1 , $0 <$ promoter score < 1 and $0 <$ coding < 1). For genes with
868 multiple scores (for example, those predicted as targets from multiple independent
869 risk signals or predicted to be impacted in several categories), we recorded the
870 highest score. Driver and transcription factor gene enrichment analysis was
871 carried out using INQUISIT scores before adding a point for driver gene status.
872 Modifications to the pipeline since original publication² included:

- 873 • Topologically associated domain boundary definitions from ENCODE T-47D
874 Hi-C analysis. Previously, we used regions from Rao, Cell 2013.
- 875 • eQTL (addition of NHS normal and tumor samples).
- 876 • Allele-specific imbalance using TCGA and Genotype-Tissue Expression RNA-
877 Seq data¹⁰⁰.
- 878 • Capture Hi-C data from six breast cell lines¹⁰³.
- Additional bio-features derived from global enrichment in this study.
- Variants affecting sites of post-translational modification¹⁰².

879 **Multi-signal targets.** To test whether more genes were targeted by multiple signals than
880 would be expected by chance, we modeled the number of signals per gene by negative
881 binomial regression (R function glm.nb; package MASS) and Poisson regression (R
882 function glm; package stats) with ChIA-PET interactions as a covariate, and adjusted
883 by fine-mapping region. Likelihood ratio tests were used to compare goodness of fit.
884 Rootograms were created using the R function rootogram (package vcd).

885 **Pathway analysis.** The pathway gene set database dated 1 September 2018 was
886 used¹⁰⁴ (http://download.baderlab.org/EM_Genesets/current_release/Human/symbol/). This database contains pathways from Reactome¹⁰⁵, the NCI Pathway
887 Interaction Database¹⁰⁶, Gene Ontology¹⁰⁷, HumanCyc¹⁰⁸, MSigdb¹⁰⁹, NetPath¹¹⁰
888 and Panther¹¹¹. All duplicated pathways, defined in two or more databases, were
889 included. To provide more biologically meaningful results, only pathways that
890 contained ≤ 200 genes were used.

891 We interrogated the pathway annotation sets with the list of high-confidence
892 (level 1) INQUISIT genes. The significance of over-representation of the
893 INQUISIT genes within each pathway was assessed with a hypergeometric test
894 using the R function phyper as follows:

$$P(x|n, m, N) = 1 - \sum_{i=0}^{x-1} \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}$$

where x is the number of level 1 genes that overlap with any of the genes in the pathway, n is the number of genes in the pathway, m is the number of level 1 genes that overlap with any of the genes in the pathway dataset ($m_{\text{strong GO}} = 145$; $m_{\text{ER-positive GO}} = 50$; $m_{\text{ER-negative GO}} = 27$; $m_{\text{ER-neutral GO}} = 73$; $m_{\text{strong pathways}} = 121$; $m_{\text{ER-positive pathways}} = 38$; $m_{\text{ER-negative pathways}} = 21$; $m_{\text{ER-neutral pathways}} = 68$) and N is the number of genes in the pathway dataset ($N_{\text{genes GO}} = 14,252$; $N_{\text{genes pathways}} = 10,915$). We only included pathways that overlapped with at least two level 1 genes. We used the Benjamini–Hochberg FDR¹¹² at the 5% level.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The credible set of causal variants (determined by either multinomial stepwise regression or PAINTOR) is provided in Supplementary Table 2c. Further information and requests for resources should be directed to M.K.B. (bcac@medschl.cam.ac.uk).

References

59. 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).
60. Gaudet, M. M. et al. Identification of a *BRCA2*-specific modifier locus at 6p24 related to breast cancer risk. *PLoS Genet.* **9**, e1003173 (2013).
61. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multiplatform method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
62. Antoniou, A. C. et al. *RAD51* 135G → C modifies breast cancer risk among *BRCA2* mutation carriers: results from a combined analysis of 19 studies. *Am. J. Hum. Genet.* **81**, 1186–1200 (2007).
63. Barnes, D. R. et al. Evaluation of association methods for analysing modifiers of disease risk in carriers of high-risk mutations. *Genet. Epidemiol.* **36**, 274–291 (2012).
64. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
65. Zhong, H. & Prentice, R. L. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **9**, 621–634 (2008).
66. 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).
67. Baran, Y. et al. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* **28**, 1359–1367 (2012).
68. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
69. Genomes Project, C. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
70. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
71. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
72. Li, Q. et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* **152**, 633–641 (2013).
73. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
74. The ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
75. Sloan, C. A. et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).
76. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
77. Stunnenberg, H. G., International Human Epigenome Consortium & Hirst, M. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* **167**, 1897 (2016).
78. Pellacani, D. et al. Analysis of normal human mammary epigenomes reveals cell-specific active enhancer states and associated transcription factor networks. *Cell Rep.* **17**, 2060–2074 (2016).
79. Cheneby, J., Gheorghe, M., Artufel, M., Mathelier, A. & Ballester, B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-Seq experiments. *Nucleic Acids Res.* **46**, D267–D275 (2018).
80. Pruitt, K. D. et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* **42**, D756–D763 (2014).
81. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
82. Wang, J. et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812 (2012).
83. Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–D115 (2016).
84. Tan, G. & Lenhard, B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).
85. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
86. Grassi, E., Zapparoli, E., Molineris, I. & Provero, P. Total binding affinity profiles of regulatory regions predict transcription factor binding and gene expression in human cells. *PLoS ONE* **10**, e0143627 (2015).
87. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
88. McLeay, R. C. & Bailey, T. L. Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010).
89. Kichaev, G. et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* **33**, 248–255 (2017).
90. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
91. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
92. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
93. Stone, E. A. & Sidow, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15**, 978–986 (2005).
94. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
95. Desmet, F. O. et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
96. Beesley, J. et al. Chromatin interactome mapping at 141 independent breast cancer risk signals.
97. Fullwood, M. J. et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
98. Corradin, O. et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).
99. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
100. Moradi Marjaneh, M. et al. High-throughput allelic expression imbalance analyses identify 14 candidate breast cancer risk genes.
101. Dixon, J. R. et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* **50**, 1388–1398 (2018).
102. Yang, Y. et al. AWESOME: a database of SNPs that affect protein post-translational modifications. *Nucleic Acids Res.* **47**, D874–D880 (2019).
103. Beesley, J. et al. Chromatin interactome mapping at 139 independent breast cancer risk signals. Preprint at bioRxiv <https://www.biorxiv.org/content/10.1101/520916v1.article-info> (2019).
104. Merico, D., Isserlin, R. & Bader, G. D. Visualizing gene-set enrichment results using the Cytoscape plug-in enrichment map. *Methods Mol. Biol.* **781**, 257–277 (2011).
105. Vastrik, I. et al. Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.* **8**, R39 (2007).
106. Schaefer, C. F. et al. PID: the Pathway Interaction Database. *Nucleic Acids Res.* **37**, D674–D679 (2009).
107. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
108. Romero, P. et al. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* **6**, R2 (2005).
109. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
110. Kandasamy, K. et al. NetPath: a public resource of curated signal transduction pathways. *Genome Biol.* **11**, R3 (2010).
111. Thomas, P. D. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
112. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300 (1995).

Acknowledgements

We thank all of the individuals who took part in these studies, as well as all of the researchers, clinicians, technicians and administrative staff who enabled this work to be carried out. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie grant agreement number 656144. Genotyping of the OncoArray was principally funded from three sources: the PERSPECTIVE project (funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie de la Science et de l'Innovation du Québec' (through Genome Québec) and the Quebec Breast Cancer Foundation); the NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) initiative and the Discovery, Biology and Risk of Inherited Variants in Breast Cancer (DRIVE) project (NIH grants U19 CA148065 and X01HG007492); and ^{Q34} Cancer Research UK (C1287/A10118 and C1287/A16563). BCAC is funded by Cancer Research UK (C1287/A16563), by the European Community's Seventh Framework Programme under grant agreement 223175 (HEALTH-F2-2009-223175) (COGS) and by the European Union's Horizon 2020 Research and Innovation Programme under grant agreements 633784 (B-CAST) and 634935 (BRIDGES). Genotyping of the iCOGS array was funded by the European Union (HEALTH-F2-2009-223175), Cancer Research UK (C1287/A10710), 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 of the GWAS data was supported in part by NIH Cancer Post-Cancer GWAS initiative grant U19 CA 148065 (DRIVE; part of the GAME-ON initiative). For a full description of funding and ^{Q35} acknowledgments, see the Supplementary Note.

Author contributions

L.Fa., H.A., J.Bee., D.R.B., J.Al., S.Ka., K.A.P., K.Mi., P.So., A.Le., M.Gh., P.D.P.P., J.C.C., M.G.C., M.K.S., R.L.M., V.N.K., J.D.E., S.L.E., A.C.A., G.C.T., J.Si., D.F.E., P.K. and A.M.D. conceived of the study idea. L.Fa., H.A., J.Bee., D.R.B., J.Al., J.D.E., S.L.E., A.C.A., G.C.T., J.Si., D.F.E., P.K. and A.M.D. developed the methodology. J.Bee., J.P.T. and M.L. provided software. L.Fa., H.A., J.Bee., D.R.B., J.Al., S.Ka., C.Tu., M.Mor. and X.J. performed a formal analysis. S.A., K.A., M.R.A., I.L.A., H.A.C., N.N.A., A.A., V.A., K.J.A., B.K.A., B.A., P.L.A., J.Az., J.Ba., R.B.B., D.B., A.B.F., J.Ben., M.B., K.B., A.M.B., C.B., W.B., N.V.B., S.E.B., B.Bo., A.B., H.Bra., H.Bre., I.B., I.W.B., A.B.W., T.B., B.Bu., S.S.B., Q.C., T.C., M.A.C., N.J.C., I.C., F.C., J.S.C., B.D.C., J.E.C., J.C., H.C., W.K.C., K.B.M., C.L.C., J.M.C., S.C., F.J.C., A.C., S.S.C., C.C., K.C., M.B.D., M.D.H., P.D., O.D., Y.C.D., G.S.D., S.M.D., T.D., I.D.S., A.D., S.D., M.Dum., M.Dur., L.D., M.Dw., D.M.E.,

C.E., M.E., D.G.E., P.A.F., U.F., O.F., G.F., H.F., L.Fo., W.D.F., E.F., L.Fr., D.E., M.Ga., M.G.D., G.Ga., P.A.G., S.M.G., J.Ga., J.A.G., M.M.G., V.G., G.G.G., G.Gl., A.K.G., M.S.G., D.E.G., A.G.N., M.H.G., M.Gr., J.Gr., A.G., P.G., E.H., C.A.H., N.H., P.Ha., U.H., P.A.H., J.M.H., M.H., W.H., C.S.H., B.A.M., J.H., P.Hi., F.B.L., A.H., M.J.H., J.L.H., A.Ho., G.H., P.J.H., E.N.I., C.I., M.I., A.Jag., M.J., A.Jak., P.J., R.J., R.C.J., E.M.J., N.J., M.E.J., A.Juk., A.Jun., R.Ka., D.K., B.Pes., R.Ke., M.J.K., E.K., J.I.K., J.K., C.M.K., Y.K., I.K., V.K., S.Ko., K.K.S., T.K., A.K., K.K., Y.L., D.L., E.L., G.L., J.Le., F.L., A.Li., W.L., J.Lo., A.Lo., J.T.L., J.Lu., R.J.M., T.M., E.M., A.Ma., M.Ma., S.Man., S.Mag., M.E.M., K.Ma., D.M., R.M., L.M., C.M., N.Me., A.Me., P.M., A.Mi., N.Mi., M.Mo., F.M., A.M.M., V.M.M., T.A., S.A.N., R.N., K.L.N., N.Z.N., H.N., P.N., F.C.N., L.N.Z., A.N., K.O., E.O., O.I.O., H.O., N.O., A.O., V.S.P., J.Pa., S.K.P., T.W.P.S., M.T.P., J.Pau., I.S.P., B.Pei., B.Y.K., P.P., J.Pe., D.P.K., K.Pr., R.P., N.P., D.P., M.A.P., K.Py., P.R., S.J.R., J.R., R.R.M., G.R., H.A.R., M.R., A.R., C.M.R., E.S., E.S.H., D.P.S., M.Sa., C.Sa., E.J.S., M.T.S., D.F.S., R.K.S., A.S., M.J.S., B.S., P.Sc., C.Sc., R.J.S., L.S., C.M.D., M.Sh., P.Sh., C.Y.S., X.S., C.F.S., T.P.S., S.S., M.C.S., J.J.S., A.B.S., J.St., D.S.L., C.Su., A.J.S., R.M.T., Y.Y.T., W.J.T., J.A.T., M.R.T., M.Te., S.H., M.B.T., A.T., M.Th., D.L.T., M.G.T., M.Ti., A.E.T., R.A.E., I.T., D.T., G.T.M., M.A.T., N.T., M.Tz., H.U.U., C.M.V., C.J.A., L.E.K., E.J.R., A.Ve., A.Vi., J.V., M.J.V., Q.W., B.W., C.R.W., J.N.W., C.W., H.W., R.W., A.W., A.H.W., D.Y., Y.Z. and W.Z. provided resources. K.Mi., J.D., M.K.B., Q.W., R.Ke., J.C.C. and M.K.S. curated and managed the data. L.Fa., H.A., J.Bee., G.C.T., D.F.E., P.K. and A.M.D. wrote the original draft of the manuscript. D.R.B., J.Al., P.So., A.Le., V.N.K., J.D.E., S.L.E., A.C.A. and J.Si. wrote and edited the manuscript. L.Fa., H.A., J.Bee. and C.Tu. visualized the results. A.C.A., G.C.T., J.Si., D.F.E., P.K. and A.M.D. supervised the project. L.Fa., P.D.P.P., J.C.C., M.G.C., M.K.S., R.L.M., V.N.K., J.D.E., S.L.E., A.C.A., G.C.T., J.Si., D.F.E., P.K. and A.M.D. acquired funding. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-019-0537-1>.

Correspondence and requests for materials should be addressed to P.K. or A.M.D.

Reprints and permissions information is available at www.nature.com/reprints.

QUERY FORM

Nature Genetics	
Manuscript ID	[Art. Id: 537]
Author	Laura Fachal

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Nature of Query
Q1:	Per journal style, separate institutes/organizations should be listed as separate affiliation addresses. Please indicate whether the address 'Genomic Medicine, Division of Evolution and Genomic Sciences, The University of Manchester, Manchester Academic Health Science Centre, Manchester Universities Foundation TrustSt. Mary's HospitalManchesterUK' should be split into two (i.e. 'Genomic Medicine, Division of Evolution and Genomic Sciences, The University of Manchester, Manchester, UK' and 'Manchester Academic Health Science Centre, Manchester Universities Foundation Trust, St. Mary's Hospital, Manchester, UK').
Q2:	Please check the order of the following affiliation addresses is correct as edited (if A is part of B and B is part of C, the order should be A, B, C (separated by commas): 'Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer InstituteNational Institutes of HealthBethesdaMDUSA' and 'University Clinic of Radiotherapy and OncologyMedical Faculty, Ss. Cyril and Methodius University in SkopjeSkopjeRepublic of North Macedonia'.
Q3:	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.
Q4:	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.
Q5:	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).
Q6:	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes.
Q7:	Please ensure that genes are correctly distinguished from gene products: for genes, official gene symbols (e.g., NCBI Gene) for the relevant species should be used and italicized; gene products such as proteins and noncoding RNAs should not be italicized.
Q8:	Your paper has been copy edited. Please review every sentence to ensure that it conveys your intended meaning; if changes are required, please provide further clarification rather than reverting to the original text. Please note that formatting (including hyphenation, Latin words, and any reference citations that might be mistaken for exponents) has been made consistent with our house style.

QUERY FORM

Nature Genetics	
Manuscript ID	[Art. Id: 537]
Author	Laura Fachal

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Nature of Query
Q9:	Please check the footnote to Table 1 carefully to ensure that all of the edits retain the intended meaning. In addition, please check the definition of 'FP' in footnote 'j' and confirm whether the definition of 'P' is required since there are no 'P' labels in the table itself. Also, please note that the part about brackets has been removed from footnote 'k' as there are no brackets around genes.
Q10:	Since the paragraphs following the sentence 'We found significant enrichment of CCVs (FDR 5%) in four genomic features...' are so long, the numerals i-iv have been replaced with level 3 headings and where required, the following sentences have been reworded for clarity. Please check the meaning is retained.
Q11:	Please check that the edits to the Fig. 3 caption retain the intended meaning. In the Fig. 3a caption, please define 'Open chrom' and 'Heterochrom'. In the Fig. 3b caption, please define 'Luminal pr'.
Q12:	In the original document, Figs 4 and 5 seemed to be in the wrong order. Please check that Figs 4 and 5, their captions and citations are all correct now.
Q13:	Please check that the edits to the Fig. 4 caption retain the intended meaning (especially the definition of 'ChIP-Seq BS'; is this correct, or does the 'BS' stand for 'binding sites?'). Also, please check the number of genes, as there appear to be 69 rather than 79 in the figure.
Q14:	Please check that the edits to the sentence 'Based on genome-wide chromosome conformation capture from...' retain the intended meaning.
Q15:	'Three examples of INQUISIT using genomic features to identify predict target genes' has been made into a heading. Please check this is correct. Also, please clarify whether it should say 'predict' or 'identify'.
Q16:	Please provide a label (with units) for the colour scale bar in Fig. 5b.
Q17:	In the Fig. 5b caption, please define: cAMP, CARM1, cGMP, EGFR, FGFR, GATA, MAPK, MET, NOTCH, PTEN, PTK6, RAS, ROBO, ROS, TGFBR and WNT.
Q18:	In the sentence 'We also found significant over-representation of additional...' please check the following have been expanded correctly: cAMP, PI3K, FGFR, EGFR and TGFBR. Also, please expand 'NOTCH', 'RAS' and 'WNT' if appropriate.
Q19:	Please check that the edits to the sentence 'However, in other signals, we also identified four coding changes...' retain the intended meaning.
Q20:	In the sentence 'For iCOGS, all known variants from the March 2010...' please define ' r^2 ' at first mention (i.e., coefficient of determination).

QUERY FORM

Nature Genetics	
Manuscript ID	[Art. Id: 537]
Author	Laura Fachal

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e-proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Nature of Query
Q21:	The meaning of the sentence 'Once a final model had been obtained for the strong...' is unclear. Please suggest alternative wording.
Q22:	Please check that the edits to the sentence 'High linkage disequilibrium between a variant...' retain the intended meaning.
Q23:	Per journal style, variables should be in non-bold italics (unless a variable is multiple letters, in which case it is non-bold/roman). Super- or subscript labels should be roman type (unless they also represent a variable, in which case they are italic). Vectors are set in bold roman font. The magnitude of a vector is in non-bold italics, as are scalar components, tensors and matrices. Please ensure your text is consistent with this throughout.
Q24:	Please check that the edits to the sentence 'Following ref. 65, we then estimated mean...' retain the intended meaning.
Q25:	In the sentence 'We conducted a search to find motif occurrences...' the meaning of '...in the genomic featured.' is unclear. Please suggest alternative wording.
Q26:	Please check the equation in the sentence 'We ran PAINTOR four times to generate four independent...' has been reproduced correctly.
Q27:	Please check that the edits to the text 'By considering functional enrichment and joint...used only OncoArray data' retain the intended meaning.
Q28:	In the sentence 'We intersected CCV positions with multiple sources of genomic information...' you cite 'Rao et al. 2014'. Please clarify which existing reference this is and add the citation number or, if this is not in the reference list, provide full reference details.
Q29:	In the sentence 'We used computational enhancer-promoter correlations...' you cite 'He et al. 2014'. Please provide the reference citation number or full reference details for this source. Also, please clarify what the 'Key Resources Table' is. Cited tables should be numbered (i.e., Table 1 or Supplementary Table 1).
Q30:	The bullet point list says 'Previously, we used regions from Rao, Cell 2013'. Should 'Cell' say 'et al'? Also, please provide the reference citation number or full reference details for this source.
Q31:	The sentence 'Genotyping of the OncoArray was principally funded from three sources...' has been edited to make the 'three sources clearer'. Please check the intended meaning is retained.
Q32:	Please check that all funders have been appropriately acknowledged and that all grant numbers are correct.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for the data collection.

Data analysis

The software used have been described in details in Online Methods section. Softwares included: IMPUTE2, MEME Suite (FIMO, AME, fasta-shuffle-letters), Meta, R (R libraries: stats, nnet, MASS, vcd, TFBSTools, MatrixRider, multinomRob, multiwaycov), Bedtools, MACs, Variant Effect Predictor, Alamut® Batch v1.6 (tools PolyPhen-2, SIFT, MAPP, MaxEntScan, Splice-Site Finder, Human Splicing Finder), PAINTOR, liftover, RMA, GlueQC,, arrayQualityMetrics v3.24.014, MACH, Local Ancestry in admixed Populations, GISTIC2.

The custom scripts used during the study are available from the corresponding author on reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The credible set of causal variants (determined by either multinomial stepwise regression and PAINTOR) is provided in Supplementary Table S2C. Further information and requests for resources should be directed to and will be fulfilled by Manjeet Bolla (bcac@medschl.cam.ac.uk)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was made. We aimed to bring together the largest possible sample size (109,900 breast cancer cases and 88,937 controls of European ancestry) with GWAS imputed up to 1000 Genomes Project Panel to study the role of genetic variants in breast cancer. The sample size included in this study (N ~ 199,000) is 2 times larger than previous breast cancer fine-mapping studies (N ~ 100,000 samples).
Data exclusions	Established protocols were used to conduct rigorous data quality control for each GWAS at the study level (more details can be found in Michailidou et al. Nature 2017 and Amos et al. Cancer Epidemiol Biomarkers Prev 2017). Imputed variants were excluded for the following reasons: (i) info score < 0.3 and (ii) minor allele frequencies (MAF) < 0.001 in both the iCOGS and OncoArray datasets. Only samples with ER status were included in the multinomial regression analyses (cases with ER status available: 29,561 iCOGS, 55,081 OncoArray). More details about the data exclusions can be found in the Online Methods section.
Replication	This was an observational study - analyses were based on all available data. All signals show no statistically significant heterogeneity between effect sizes estimated in iCOGS and OncoArray cohorts.
Randomization	Not relevant because the study is not experimental.
Blinding	Not relevant because the study is not experimental.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Analyses were conducted on breast cancer cases and controls of European ancestry. The association between each variant and breast cancer risk was adjusted by study (iCOGS) or country (OncoArray), and eight (iCOGS) or ten (OncoArray) ancestry-informative principal components.
Recruitment	Epidemiological data for European women were obtained from 75 breast cancer case-control studies participating in the Breast Cancer Association Consortium (BCAC). The majority of studies are population-based case-control studies, or case-control studies nested within population-based cohorts, but a subset of studies oversampled cases with a family history of the disease. Subjects included from CIMBA are women of European ancestry aged 18 years or older with a pathogenic variant for BRCA1. The majority of the participants were sampled through cancer genetics clinics. Multiple members of the same family were included in some instances.
Ethics oversight	All participating studies were approved by their appropriate ethics review board and all subjects provided informed consent.

Note that full information on the approval of the study protocol must also be provided in the manuscript.