A case-only study to identify genetic modifiers of breast cancer risk for *BRCA1/B RCA2* mutation carriers

Breast cancer (BC) risk for *BRCA1* and *BRCA2* mutation carriers varies by genetic and familial factors. About 50 common variants have been shown to modify BC risk for mutation carriers. All but three, were identified in general population studies. Other mutation carrier-specific susceptibility variants may exist but studies of mutation carriers have so far been under-powered. We conduct a novel case-only genome-wide association study comparing genotype frequencies between 60,212 general population BC cases and 13,007 cases with *BRCA1* or *BRCA2* mutations. We identify robust novel associations for 2 variants with BC for *BRCA1* and 3 for *BRCA2* mutation carriers, *P* < 10^-8, at 5 loci, which are not associated with risk in the general population. They include rs60882887 at 11p11.2 where *MADD*, *SP11* and *EIF1*, genes previously implicated in BC biology, are predicted as potential targets. These findings will contribute towards customising BC polygenic risk scores for *BRCA1* and *BRCA2* mutation carriers.
Breast cancer (BC) is the most common cancer in women worldwide and BC family history is one of the most important risk factors for the disease. Women with a history of BC in a first-degree relative are about two times more likely to develop BC than women without a family history. Around 15–20% of the familial risk of BC can be explained by rare mutations in the BRCA1 or BRCA2 genes. A recent prospective cohort study estimated the cumulative risk of BC by 80 years to be 72% for BRCA1 mutation carriers and 69% for BRCA2 mutation carriers. This study also demonstrated that BC risk for mutation carriers varies by family history of BC in first and second degree relatives, suggesting the existence of other genetic factors that modify BC risks.

A total of 179 common BC susceptibility single nucleotide polymorphisms (SNPs) or small insertions or deletions (INDELs) have been identified through genome-wide association studies (GWAS) in the general population. Although risk alleles at individual SNPs (hereafter used as a generic term to refer to common variants, which also includes the small INDELs) are associated with modest increases in BC risk, it has been shown that they combine multiplicatively on risk, resulting in substantial levels of BC risk stratification in the population. Similarly, more than 50 of the common genetic BC susceptibility variants have also been shown to be associated with BC for BRCA1 and BRCA2 mutation carriers and their joint effects, summarised as polygenic risk scores (PRS), result in large differences in the absolute risks of developing BC for mutation carriers at the extremes of the PRS distribution. BC GWAS for BRCA1 and BRCA2 mutation carriers have been carried out through the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). However, despite the large number of BRCA1 and BRCA2 mutation carriers included, the power to detect genetic modifiers of risk remains limited in comparison to that available in the general population. To date, no variants specifically associated with BC risk for BRCA1 and BRCA2 carriers have been identified.

Here, we apply a novel strategy using a case-only GWAS design, in which SNP genotype frequencies in 7,257 BRCA1 and 5,097 BRCA2 mutation carrier BC cases are compared to those in 60,212 BC cases from the Breast Cancer Association Consortium (BCAC), unselected for mutation status. We aim (1) to identify novel SNPs that modify BC risk for BRCA1 or BRCA2 mutation carriers but are not associated with risk in the general population and (2) for the known 179 BC susceptibility SNPs, assess whether there is evidence of an interaction between the SNPs and BRCA1 or BRCA2 mutations and therefore evaluate whether the SNP effect size estimates applicable to mutation carriers are different.

We identify robust novel associations for 2 variants with BC for BRCA1 and 3 for BRCA2 mutation carriers, \( P < 10^{-8} \), at 5 loci, which are not associated with risk in the general population. They include rs60882887 in 11p11.2 where MADD, SP11 and EIF1, genes previously implicated in BC biology, are predicted as potential targets. These findings will contribute towards customising BC PRS for BRCA1 and BRCA2 mutation carriers.

### Results

#### Sample characteristics

A total of 60,212 BCAC cases and 7,257 BRCA1 mutation carrier cases were available for the BRCA1 case-only analyses and 57,725 BCAC cases and 5,097 BRCA2 mutation carrier cases were available for the BRCA2 case-only analyses (Fig. 1). A total of 45,881 BCAC controls and 5,750 unaffected BRCA1 mutation carriers were available for the BRCA1 control-only analyses and 43,549 BCAC controls and 4,456 unaffected BRCA2 mutation carriers for the BRCA2 control-only analyses (see Fig. 2). Only women of European ancestry were included with 60.9% samples from European countries, 31.1% from the USA, 6.1% from Australia and 1.7% from Israel (Supplementary Tables 1–4). The mean age at BC diagnosis for mutation carrier cases in CIMBA was 42.5 years (40.9 for BRCA1 mutation carriers; 44.1 for BRCA2 mutation carriers) and 58.4 years for cases in BCAC.

The analytical process for assessing interactions with known BC susceptibility SNP is summarised in Fig. 3 and for the detection of novel modifiers in Fig. 4.

#### Independence of SNP frequency with mutation carrier status

Under a case-only study design, it is important to establish independence between the SNPs and BRCA1 or BRCA2 mutation carrier status. This was assessed a genome-wide level using a control-only analysis which included controls from BCAC and unaffected mutation carriers from CIMBA with SNP data imputed based on the 1,000 genomes project. Genotypes had been

---

**Fig. 1 Case-only sample selection.** Sample selection for a BRCA1 and b BRCA2 case-only analysis. *Four studies were excluded because they were included in clinical trials based on breast tumour characteristics as HER-2 receptor status (see Supplementary Table 2).**
Imputed separately by each consortium. In the analysis of BRCA1 mutation carriers, 2,164 SNPs were excluded because they were located in or within 500 kb of BRCA1. 2,070 SNPs were excluded from further analyses because they showed associations at $p < 10^{-8}$ with BRCA1 mutation carrier status in the control-only analysis (2,012 SNPs located on chromosome 17 and 58 on other chromosomes). In the analysis of BRCA2 mutation carriers, 2,947 SNPs were excluded because they were located in or within 500 kb of BRCA2. A further 626 SNPs were excluded from further analyses because they were found to be associated with BRCA2 mutation carrier status in the control-only analysis (566 SNPs on chromosome 13, and 60 on other chromosomes). A total of 9,068,301 SNPs remained for the BRCA1 case-only association analysis and 9,043,830 SNPs for the BRCA2 case-only analysis.

Interactions with known BC susceptibility SNPs. Based on published data, 179 SNPs were considered as established BC susceptibility SNPs (Fig. 3); 158 SNPs were associated with overall BC risk and 21 additional SNPs were found to be associated through studies in ER-negative breast cancer (see Supplementary Table 11 in Milne et al. One of the 158 SNPs, rs11571833 located within BRCA2 was excluded from the BRCA2 analysis. The detailed results are shown in Supplementary Data 1–3.

For BRCA1 mutation carriers, previous studies have demonstrated heterogeneity in the associations of the SNPs with ER-positive and ER-negative breast cancer. Since BRCA1 mutation carriers develop primarily ER-negative BC, to comprehensively assess the evidence of interaction with BRCA1 mutation status, we followed a two-step process; we first assessed the associations using all BC cases from BCAC and then we restricted the comparison to BCAC ER-negative BC cases. Of the 158 SNPs, 59 were associated with BRCA1 mutation carrier status when compared to all BC cases ($P < 0.05$, Supplementary Data 1). However, after adjusting for multiple testing, only four of these SNPs were significantly associated with BRCA1 mutation carrier status.
SNPs were associated ($P < 2.7 \times 10^{-4}$) and also showed evidence of association ($P < 0.05$) when compared with ER-negative BC cases (Table 1). Two additional SNPs on chromosome 1 and 6 (chr1_10566215_A_G and rs17529111) were associated at $P < 2.7 \times 10^{-4}$ with BRCA1 mutation status only when compared with ER-negative BCAC cases. The OR estimates for association with BRCA1 mutation status for these six SNPs were similar under both case-only analyses (all BC and ER-negative BC cases analyses) and varied from 0.85 to 1.07, suggesting that the magnitude of their associations with BC risk for BRCA1 mutation carriers differs from that observed in the general population. For the other 152 SNPs, there was no evidence of association with BRCA1 mutation status for these six SNPs were similar under both case-only analyses (all BC and ER-negative BC cases analyses) and varied from 0.85 to 1.07, suggesting that the magnitude of their associations with BC risk for BRCA1 mutation carriers differs from that observed in the general population. For the other 152 SNPs, there was no evidence of association with BRCA1 mutation status when compared against the ER-negative BCAC cases from BCAC (Supplementary Data 1), suggesting that the OR estimated using case-control data from BCAC was greater than that in the general population (ORBCAC) and for two of these three, the ORcomputed was in the opposite direction than the ORBCAC (Table 1). For the four other SNPs (rs13281615, chr16_52599188_C_T, chr1_10566215_A_G and rs66823261), the estimated interaction OR resulted in the OR for associations with BRCA1 BC risk being closer to 1 (Table 1).

Among the remaining 172 SNPs (152 + 20) that showed no associations with BRCA1 mutation status, the estimated ORcomputed was smaller (i.e., closer to 1) than those estimated in the general population (ORBCAC) for 146 SNPs (85%,...
Novel SNP modifiers. To identify novel SNPs that modify BC risks for BRCA1 and BRCA2 mutation carriers, we used a case-only design to investigate the associations of SNPs that had not been previously shown to be associated with BC in the general population (Fig. 4).

For BRCA1 mutation carriers, a total of 924 SNPs showed associations at $P < 10^{-8}$ in all BC case-only analysis. To ensure that none of these associations are driven by differences in the distribution of ER-positive and ER-negative tumours in BCAC cases, an intermediate step was applied, in which we re-analysed the associations after restricting the BCAC data to only ER-negative cases. 220 of these SNPs remained significant at $P < 10^{-7}$ located in 11 distinct genomic regions. SNPs were considered to belong to the same region if they were located within 500 kb of each other.

To ensure that none of these associations were driven by differences in the genotype imputation in the BCAC and CIMBA data (which had been carried out separately), all the SNPs in these 11 distinct genomic regions were re-imputed in the BCAC and CIMBA samples jointly and the associations for all SNPs in the regions were re-assessed in the control-only and case-only analyses. After the exclusion of 614 SNPs (613 on chromosome 17) that showed associations in the control-only analysis, 71 SNPs in two regions remained significant at $P < 10^{-8}$ (Supplementary Data 4) in the case-only analyses including all BCAC cases. None of these SNPs had been previously reported in GWAS in the general population ($p$-values of association ranged from 0.51 to 5.9 x 10^{-5} with effect sizes in the range 0.96–1.04 in BCAC case-control analyses)\(^{35,48}\). A forward step-wise regression analysis within each of these two regions (restricted to the SNPs exhibiting associations at $P < 10^{-8}$) starting with the most significant SNP and adding sequentially the other SNPs, identified a set of four conditionally independent SNPs (top SNPs) (Table 3): all SNPs were imputed, with $r^2 > 0.5$, and had minor allele frequency (MAF) > 10%. Three of the top SNPs are located in 17q21.2. rs58117746 is an insertion of 16 bp within an exon of KRTAP4-5 leading to a frameshift of the amino acid sequence. rs5820435 and rs11079012 are both intronic and located in LEFRA4 (also named P3H4) and JUP, respectively, while rs80221606 is intronic and located in 11p11.2, within CELFI1. The OR estimates of these four top SNPs ranged from 0.78 to 1.22. All showed evidence of heterogeneity in the OR by country ($P < 0.05$) (Table 3); however, in a leave-one-out analysis, in which each country was left out in turn, the overall

Table 1. Known BC susceptibility SNPs demonstrating associations in the BRCA1 case-only analysis.

| Location | SNP name | Chr Position | Nearest gene | Estimated effect allele | Ref Allele | Frequency | OR BCAC = 1.04 | OR BCAC > 1.0 | OR estimated by country |
|----------|----------|--------------|--------------|-------------------------|-----------|-----------|----------------|---------------------|------------------------|
| 10q22.3  | rs1543504 | 10,657,110   | ZM1          | T                       | C         | 0.04      | 1.01           | 1.06                | 1.00                   |
| 1p36.22  | rs199231  | 1,056,621    | RPL23AP53    | C                       | T         | 0.05      | 1.03           | 1.03                | 1.00                   |
| 8p23.3   | rs6682326 | 17,069,214   | RPL23AP53    | T                       | C         | 0.04      | 1.04           | 1.04                | 1.00                   |

All BC SNPs associated with all BC in the general population. N = 47,410 breast cancer cases (6,228 BCAC cases and 41,182 BCAC mutation carrier cases). 

Of the 154 SNPs that showed no significant associations with BRCA2 mutation status, 79% had ORs of BC for BRCA2 mutation carriers ($\text{OR}_{\text{bcAC}}$) that were closer to 1 when compared to the ORs estimated using data in the general population ($\text{OR}_{\text{bcAC}}$) (Supplementary Data 3).
associations remained similar (Supplementary Fig. 1 and 2) suggesting that no individual country had a big impact on the observed associations.

For BRCA2 mutation carriers, the case-only analysis identified 273 SNPs, located across 22 regions, with evidence of association at \( P < 10^{-8} \). After the joint re-imputation of the SNPs in these 22 regions, only 102 SNPs located in four regions (2p14, 13q13.1, and 13q13.2) remained associated at \( P < 10^{-8} \) (Supplementary Data 5). The step-wise regression analysis suggested that associations in each of the four regions were driven by a single variant (top SNPs) (Table 4). All four variants were imputed with \( r^2 > 0.5 \) and had MAF higher than 5%. At 2p14, rs12470785 \( (r^2 = 0.98) \) is within an intron of ETFA1. At 13q13.1, rs79183898 \( (r^2 = 0.84) \) is located between B3GALT1 and RXFP2 and rs736596 \( (r^2 = 0.66) \) is within an intron of STARD13. At 13q13.2, rs4943263 \( (r^2 = 0.99) \) is located between RPI1-266E6.3 and RPI1-307013.1. None of these SNPs had been previously reported to be associated with BC risk in BCAC studies in the general population. \( p \)-values from 0.01 to 0.90 in BCAC case-control analyses. The OR estimates of these four SNPs ranged from 0.85 to 1.37. All showed evidence of heterogeneity in the OR by country at \( p = 0.05 \) (Table 4). In the leave-one-country-out sensitivity analysis the two intergenic SNPs, rs79183898 and rs736596 were no longer significant at \( P < 10^{-4} \) when studies from the USA were excluded from the analysis and the OR estimates were substantially attenuated (Supplementary Figs. 3 and 4).

### In silico analyses on credible causal variants (CCV).

In order to determine the likely target genes of each region of the eight novel mutation carriers’ BC risk-associated SNPs, we first defined credible set of SNPs candidates to be causal (credible causal variants [CCVs]) (see “Methods”).

Sets of CCVs were sought for the two regions found in the previous step-wise analyses to be associated with risk in BRCA1 mutation carriers. In the region located at 11p11.2, only one signal composed of 74 CCVs was found (Table 5). All these 74 CCVs were imputed with a \( r^2 \) higher than 0.92 (Supplementary Data 6). In the region located in 17q21.2, we found nine signals which contained from one to 13 CCVs (Table 5). Two of these CCVs were genotyped and the others had an \( r^2 \) between 0.50 and 0.98 (Supplementary Data 6).

We used INQUISIT to prioritize target genes by intersecting each CCV with publicly available annotation data from breast cells and tissues (see “Methods”). The results for BRCA1 mutation carriers are summarized in Supplementary Data 7. For BRCA1 mutation carriers, we predicted 38 unique target genes for six of the 10 independent signals. Seven target genes in two regions (MTCH2, MADD, PSMC3, RP11-750H9.5, SLC39A13, SPI1, and EIF1) were predicted with high confidence (designated Level 1, scoring range between Level 1 [highest confidence] to Level 3 [lowest confidence]). All seven Level 1 genes were predicted to be distally regulated by CCVs.

Similarly, sets of CCVs were sought from the four regions found in the previous step-wise analyses to be associated with risk in BRCA2 mutation carriers. A total of 17 signals were found. One signal composed of 78 CCVs was found in the region located at 2p14 (Table 6). One CCV was genotyped and the others were imputed with \( r^2 \) between 0.95 and 0.99 (Supplementary Data 8). Twelve signals were found from the two regions previously found in 13q13.1 which contained from one to 46 CCVs. The analysis in the region of rs79183898 in 13q13.1 found three signals out of the 12, which are located in 13q13.2 (with top SNPs: rs71434801, rs77197167, rs114300732). Finally, four signals in the previously identified region located in 13q13.2 containing from three to 40 CCVs were found. Among all CCVs, 11 are genotyped and the imputed ones have an \( r^2 \) higher than 0.58 (Table 6 and Supplementary Data 8). For BRCA2 mutation carriers, we predicted 24 unique target genes for 10 of the 17 independent signals, including one high confidence target gene, \( STARID3 \) (Table 6 and Supplementary Data 8).

### Discussion

To identify novel genetic modifiers of BC risk for BRCA1 and BRCA2 mutation carriers and to further clarify the effects of known BC susceptibility SNPs on BC risk for carriers, a novel case-only analysis strategy was used based on GWAS data from unselected BC cases in BCAC and mutation carriers with BC from CIMBA. This strategy provides increased statistical power for detecting new associations and for clarifying the risk associations of known BC susceptibility SNPs in mutation carriers.

Of the 179 known BC susceptibility SNPs identified through GWAS in the general population, only 10 showed evidence of interaction with BRCA1 or BRCA2 mutation carrier status after taking the tumour ER-status into account. None of these 10 SNPs was among the fifty SNPs previously shown to be associated with BC for mutation carriers. However, 82% of all 179 known susceptibility SNPs showed a predicted OR point estimate for mutation carriers closer to 1 than that estimated in the general population. The effect sizes in the general population may be

| Location | SNP name | Chr | Position | Nearest gene | Estimated effect allele | Referent allele | Frequency | OR | p | ORBCAC | \( P_{BCAC} \) | \( OR_{BCAC} \) | Variation in risk |
|----------|----------|-----|----------|--------------|------------------------|---------------|-----------|----|----|---------|------------|---------|----------------|
| 5q11.2   | rs62355902 | 5   | 54063733  | MAPK1       | A                      | A              | 0.18       | 0.98 | 0.89 | 1.00e-24 | 1.18       | 8.50e-42 |             |
| 9q31.1   | rs10579243 | 9   | 11030615  | RP11-4389P9  | A                      | C              | 0.31       | 1    | 0.89 | 4.60e-08 | 1.06       | 4.20e-10 |             |
| 22q13.1  | chr22:40876234 | 22 | 40876234  | MKI1       | C                      | T              | 0.11       | 1    | 0.88 | 2.80e-04 | 1.12       | 5.70e-16 | 0.98         |

\( N = 62,822 \) breast cancer cases (57,725 BCAC cases and 5,097 BRCA2 mutation carrier cases).

Considering SNPs with known BC (Michailidou et al.) associations in the general population.

T77 tends to 1.0D increase in same direction, 0.0D increase in opposite direction.

*After allowing for multiple testing, \( \alpha = 2.7 \times 10^{-4} \).
Table 3 List of potential novel SNP modifiers associated in the case-only analysis for BRCA1 mutation carriers.

| Location | SNP namea | Chrb | Positionc | Nearest gene | Localization | Estimated effect allele | Reference allele | p-d | Frequencyd | ORf | Pd | HR(CIMBA) | PCIMBA | ORBCACj | PBCACj | Target geneo |
|----------|-----------|------|-----------|--------------|--------------|------------------------|-----------------|-----|-------------|-----|----|-----------|--------|----------|--------|--------------|
| 1q11.2   | rs80221606| 11   | 47560211  | CELF1        | Intronic     | A                      | A               | 0.76| 0.10        | 0.78| 1.12e-10 | 0.76      | 6.36e-07 | 0.98     | 7.60e-01     | 104 | 0.01 | 1.39e-03 | 1.51e-03 | Level 2   |
| 1q11.2   | rs81371746| 17   | 39935775  | KRTAP4-5     | Intronic     | T                      | T               | 0.50| 0.39        | 0.82| 2.86e-07 | 0.82      | 4.95e-01 | 0.88     | 7.77e-05     | 101 | 0.07 | 1.06e-08 | 1.06e-08 | -        |
| 1q11.2   | rs5820435 | 17   | 39916558  | LILPER4      | Intronic     | C                      | C               | 0.66| 0.31        | 1.17| 7.06e-06 | 0.18      | 2.35e-06 | 0.98     | 3.10e-01     | 101 | 0.31 | 1.13e-07 | 1.13e-07 | Level 2   |

N = 67,469 breast cancer cases (60,212 BCAC cases and 7,257 BRCA1 mutation carrier cases).

*The most significant SNP of each region after allowing for multiple testing, α* = 10^-8.

aChromosome.
bBuild 37 position.
cBuild 38 position.
dImputation accuracy.
eFrequency of the allele for which effect is estimated in BCAC cases (OncoArray dataset).
fPer allele odds ratio estimated in the case-only analysis. OR values were computed from a two sided logistic regression using a 1 df test adjusted for age at BC diagnosis, country and the first four principal components.
gp-value in the case-only analysis.
hPer allele odds-ratio estimated in the case-only ER-negative subgroup analysis.
iP-value found in CIMBA cohort analysis.
jPer allele hazard ratio estimated in CIMBA cohort analysis.
kP-value of the heterogeneity test by country.
lINQUISIT score level: 1 = most functional evidence supporting a potential link between CCVs and target gene.

Table 4 List of potential novel SNP modifiers associated in the case-only analysis for BRCA2 mutation carriers.

| Location | SNP namea | Chrb | Positionc | Nearest gene | Localization | Estimated effect allele | Reference allele | p-d | Frequencyd | ORf | Pd | HR(CIMBA) | PCIMBA | ORBCACj | PBCACj | Target geneo |
|----------|-----------|------|-----------|--------------|--------------|------------------------|-----------------|-----|-------------|-----|----|-----------|--------|----------|--------|--------------|
| 2p14     | rs12470785| 2    | 67634003  | ETAA1        | Intronic     | A                      | A               | 0.98| 0.30        | 0.84| 2.38e-11 | 0.84      | 1.69e-05 | 0.98   | 0.03 | 2.18e-07 | 1.51e-03 | Level 2   |
| 13q11.1  | rs9783898 | 13   | 32221794  | B3GALTL - RXFP2 | Intergenic | A                      | T               | 0.84| 0.39        | 1.17| 3.34e-03 | 0.94      | 1.01e-12 | 0.94   | 1.12e-07 | 0.54 | 1.12e-08 | 1.00e-08 | Level 1   |
| 13q11.1  | rs736596  | 13   | 33776506  | STARD13      | Intronic     | G                      | C               | 0.99| 0.27        | 1.17| 8.33e-10 | 0.98      | 9.83e-03 | 0.98   | 0.45 | 4.99e-11 | 0.49e-03 | Level 2   |
| 13q13.2  | rs4943263 | 13   | 35736357  | RP11-266656.3 - RP11-3070131 | Intergenic | T                      | T               | 0.99| 0.27        | 1.17| 8.33e-10 | 0.98      | 9.83e-03 | 0.98   | 0.45 | 4.99e-11 | 0.49e-03 | Level 2   |

N = 62,822 breast cancer cases (57,725 BCAC cases and 5,097 BRCA2 mutation carrier cases).

*The most significant SNP of each region after allowing for multiple testing, α* = 10^-8.

aChromosome.
bBuild 37 position.
cBuild 38 position.
dImputation accuracy.
eFrequency of the allele for which effect is estimated in BCAC cases (OncoArray dataset).
fPer allele odds ratio estimated in the case-only analysis. OR values were computed from a two sided logistic regression using a 1 df test adjusted for age at BC diagnosis, country and the first four principal components.
gp-value in the case-only analysis.
hPer allele odds-ratio estimated in BCAC (Michailidou et al. 2017). For SNPs with PH>1 0^-3, significance was attained in merging data of Oncoarray, iCOGS and 11 different breast cancer GWAS in Michailidou et al. 2017.

iP-value found in CIMBA cohort analysis.

jPer allele hazard ratio estimated in CIMBA cohort analysis.
kP-value of the heterogeneity test by country.
lINQUISIT score level: 1 = most functional evidence supporting a potential link between CCVs and target gene.
### Table 5: List of most significant SNPs in the CCV analysis for BRCA1 mutation carriers.

| Fine mapping region | Signal | Chr² | Location | SNP name⁴ | SNP with p-value < 5.00e⁻¹⁰ | Nearest gene | Localisation | Estimated effect allele | Referent allele | Frequency | ORi | ORi | ORi | PCEM,⁶ | HRCMAI²⁶ |
|---------------------|--------|------|----------|-----------|---------------------------|--------------|--------------|---------------------|----------------|-----------|-----|-----|-----|--------|--------|
| chr1:14:6737616-47773656 | 1 | 1 | 10p11.22 | rs60882887 | 0.45 | 2.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 2 | 2 | 17q21.2 | rs58204135 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 3 | 3 | 17q21.2 | rs99019133 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 4 | 4 | 17q21.2 | rs98917746 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 5 | 5 | 17q21.2 | rs58204135 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 6 | 6 | 17q21.2 | rs58204135 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 7 | 7 | 17q21.2 | rs99019133 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 8 | 8 | 17q21.2 | rs98917746 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |
| chr1:17:50638515-40148816 | 9 | 9 | 17q21.2 | rs58204135 | 0.45 | 0.95 | 1.10 | 0.99 | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 | 0.99 |

### Table 6: List of most significant SNPs in the CCV analysis for BRCA2 mutation carriers.

| Fine mapping region | Signal | Chr² | Location | SNP name⁴ | SNP with p-value < 5.00e⁻¹⁰ | Nearest gene | Localisation | Estimated effect allele | Referent allele | Frequency | ORi | ORi | ORi | PCEM,⁶ | HRCMAI²⁶ |
|---------------------|--------|------|----------|-----------|---------------------------|--------------|--------------|---------------------|----------------|-----------|-----|-----|-----|--------|--------|
| chr2:267094966-68009466 | 1 | 1 | 2p14 | rs12470785 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 2 | 2 | 13q21.3 | rs97789189 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 3 | 3 | 13q21.3 | rs77491401 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 4 | 4 | 13q21.3 | rs77491401 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 5 | 5 | 13q21.3 | rs77491401 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 6 | 6 | 13q21.3 | rs77491401 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 7 | 7 | 13q21.3 | rs77491401 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 8 | 8 | 13q21.3 | rs77491401 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |
| chr13:33015494-23354949 | 9 | 9 | 13q21.3 | rs77491401 | 0.30 | 4.00e⁻¹⁰ | 2.80 | 0.98 | 9.00e⁻⁰⁴ | 0.99 | 1.30 | 1.30 | 0.99 | 0.99 | 0.99 |

---

N = 67,469 breast cancer cases (60,212 BCAC cases and 7,257 BRCA1 mutation carrier cases).

1. Significant region in the main analysis used to look for credible causal variants (CCV).
2. Significant signal (the first one corresponds to the CCV set without any adjustment and the following are those with adjustment on each most significant SNP of the previous signals).
3. Number of credible causal variants at each signal (SNP with p-value < 2 order of magnitude of the most significant one).
4. The most significant SNP after adjustment on the most significant SNPs of the previous signals (except for those of the signal 1).
5. Frequency of the allele for which effect is estimated in BCAC cases (GoroArray dataset).
6. Per allele odds ratio estimated in the case-only analysis. OR values were computed from a two sided logistic regression using a 1df lrtest adjusted for age at BC diagnosis, country, the first four principal components and the most significant SNPs of the previous signals (except for those of the signal 1).
7. P-value in the case-only analysis restricted to ER-negative BCAC cases and after adjustment with the most significant SNP of the previous signals (except for those of the signal 1).
8. Per allele odds ratio estimated in the case-only analysis restricted to ER-negative BCAC cases and after adjustment with the most significant SNP of the previous signals (except for those of the signal 1).
9. Per allele hazard ratio estimated in CIMBA cohort analysis.
10. Per allele hazard ratio estimated in CIMBA cohort analysis.
somewhat exaggerated as the BCAC dataset used here contributed to the discovery of most of the loci, although this effect is likely to be small as most loci are highly significant and the effects have been replicated in independent datasets\(^2\). Taken together, these results suggest that, while most SNPs associated with risk in the general population are associated with risk for mutation carriers, the average effect sizes for mutation carriers are smaller. These findings are in line with previous results by Kuchenbaecker et al.\(^3\) and suggest that a PRS built using data from the general population will have a smaller effect size for BRCA1/2 mutation carriers.

For 10 SNPs, an interaction was observed with BRCA1 or BRCA2 mutation carrier status, suggesting that these SNPs have different effect sizes in BRCA1 or BRCA2 mutation carriers compared to the general population (seven for BRCA1 mutation carriers and three for BRCA2 mutation carriers). Specifically, for seven SNPs the confidence intervals were consistent with no effect on BC risk for mutation carriers, one SNP was associated with a larger OR for mutation carriers compared to the general population and two were associated in the opposite direction to that observed in the general population. However, distinguishing between a smaller effect size for mutation carriers compared to the general population OR estimates and no association for mutation carriers is very challenging since, even with the large sample size here, it is not possible to estimate precisely the effect sizes for individual variants. Larger sample sizes will be required for this purpose. Determining the precise effects of the SNPs in BRCA1 and BRCA2 mutation carriers will provide insights for understanding the biological basis of cancer development associated with BRCA1 and BRCA2 mutations.

We also identified eight novel conditionally independent common SNPs associated with BC risk (four for BRCA1 mutation carriers, four for BRCA2 mutation carriers). These have not been reported in previous association studies\(^5\)\(^6\)\(^15\)\(^18\)\(^20\)\(^39\)\(^47\). The case-only OR estimates for these SNPs varied from 0.85 to 1.37 for BRCA2 mutation carriers and from 0.78 to 1.22 for BRCA1 mutation carriers. For five of these SNPs the estimated ORs from the case-only analysis results were in the same direction as the estimated HRs from previously reported GWAS using cohort analyses restricted in BRCA1 and BRCA2 mutation carriers in CIMBA\(^56\). Two of these five SNPs also demonstrated some evidence of association in mutation carriers (\(p = 2.2 \times 10^{-2}\) for rs58117746 for BRCA1 mutation carriers; and \(p = 7.7 \times 10^{-5}\) for rs12470785 in ETAA1 for BRCA2 mutation carriers; Tables 3 and 4). For the remaining three variants, rs5820435 and rs11079012 at 17q21.2 and rs736596 at 13q13.1, the associations in BRCA1 or BRCA2 mutation carriers in the CIMBA data were not consistent with the observed interactions and might be artefactual. One possibility is that the associations with SNPs on 17q and 13q in BRCA1 and BRCA2 carriers respectively, reflect confounding due to linkage disequilibrium (LD) with specific mutations. Although we excluded variants with evidence of association in the control only analyses, it is possible that residual confounding due to specific mutations was still present.

Seven genes at a locus at 11p11.2 marked by rs60882887, were predicted with high confidence as targets, including MADD, SP11 and EIF1 which have previously been reported to be associated with BC biology\(^57\)\(^58\)\(^59\). However, no likely target genes were predicted at the 17q21.2 region. The lack of target gene predictions may be due to reliance on breast cell line data which does not represent the in vivo tissue of interest or due to the fact that the target transcripts are not annotated.

Only one gene, STARD13, was predicted as a potential target of SNPs at 13q13.1. This tumour suppressor gene has been previously implicated in metastasis, cell proliferation and development of BC\(^60\). However, rs736596, localized at 13q13.1, showed no association in CIMBA analyses and the association observed in our case-only analysis showed heterogeneity by country.

At the 2p14 locus, INQUISIT-predicted target genes included ETAA1 with lower confidence. The OR estimates obtained in the case-only analysis for the SNPs located in this gene were consistent with the HR estimated in previously reported CIMBA analyses\(^45\). Moreover, around one hundred correlated SNPs, were associated with BRCA2 mutation carrier status at \(p < 10^{-8}\), including the genotyped SNP chr2_67654113_C_T.

The validity of the case-only analysis as evidence of interaction relies on the assumption of independence between the mutation status and the SNPs under investigation\(^31\). Therefore, based on the control-only analyses, we excluded ~2,000 SNPs which were associated with BRCA1 or BRCA2 mutation carrier status and also showed an association with risk in the case-only analyses (Supplementary Fig. 5). While most of these associations are probably spurious, due to (intra- or inter-chromosomal) LD with BRCA1 or BRCA2 mutations, it is possible that some may reflect true associations and that the higher frequency in unaffected BRCA1/2 may be because they are relatives of BC cases. These associations may warrant further evaluation using other study designs. A recent publication using data from the Framingham Heart Study suggested that interchromosomal LD can be caused by bio- genetic mechanisms possibly associated with favourable or unfavourable epistatic evolution\(^52\). SNPs for which no association with mutation carrier status was found at the significance level of \(10^{-8}\) were assumed to be independent of the mutation status. However, this does not necessarily rule out residual LD between the novel SNPs on chromosomes 13 and 17 and BRCA1 or BRCA2 mutations. Therefore, the OR estimates for these SNPs might be biased and may further explain the lack of evidence of association in the CIMBA only analyses.

Our findings highlight the importance of imputation in GWAS. The imputed genome-wide genotype data used in the main case-only association analyses were based on carrying out the imputation separately for the BC cases from BCAC and CIMBA. We found that 28 out of the 33 regions associated with BRCA1 or BRCA2 mutation carrier status were no longer associated with risk after re-imputing all samples together. By re-imputing all the data together we ensured that the associations observed for the remaining regions are robust to potential differences in the imputation accuracy between the BCAC and CIMBA samples.

Under our analytical strategy, only the regions for which evidence of associated with BC risk was observed were re-imputed using all BCAC and CIMBA samples combined. This re-imputation was not done at genome-wide level due to computational constraints and this may have led to false-negative associations being excluded for further evaluation as potential novel modifiers. Future analyses should aim to analyse the genome-wide associations after the genome-wide re-imputation across the combined BCAC and CIMBA dataset. However, our approach using joint one-step imputation should have ensured that associations we report (all of which are common SNPs with imputation scores > 0.5) are not driven by inaccuracies in imputation.

Due to the recruitment of participants in CIMBA studies primarily through genetic counselling, the mean age at diagnosis of mutation carriers was 16 years younger than the BC cases participating in BCAC. Although all analyses were adjusted for age, the observed associations might be related to the ageing process instead of interactions with mutation carrier status. Another source of bias could be related to the fact that there are 1.5 times more prevalent cases among CIMBA (68.1%) than BCAC (42.3%) with a delay between diagnosis and study recruitment of 6.83 years and 2.07 years respectively. An observed association might be due to a differential survival between CIMBA and BCAC cases.
However, none of the identified SNPs has been found to be associated with BC survival.

The majority (92.5%) of cases and controls in BCAC were not tested for BRCA1/2 mutations at the time of enrolment, potentially leading to some attenuation in the interaction OR (as some BCAC cases will be carriers). However, most BCAC studies were population-based case-control studies and the proportion of cases and controls that carry pathogenic BRCA1/2 mutations will be small (<5%), hence any attenuation is likely to be negligible.

Despite heterogeneity in the interaction ORs by country for some SNPs, results were generally robust to the exclusion of each study sequentially except, for two SNPs (rs79183898 and rs736596) found associated with BRCA2 mutation carrier status; for these, the association seemed to be driven by data from the USA. For the other SNPs, the observed heterogeneity may be due to random error, given the relatively small sample sizes of each country. However, if these differences are real, future PRs for BRCA1 and BRCA2 carriers should consider the country-specific differences.

This is the first analysis of genetic modifiers of BC risk that investigated the differences in the association of common genetic variants with BC risk in the general population and in women with BRCA1 or BRCA2 mutations. The inclusion of unselected BC cases resulted in increased sample size and hence a gain in statistical power for identifying novel SNPs. These represent the largest currently available datasets, but is important to replicate these observations in independent samples. This should be possible through the ongoing CONFLUENCE (https://dceg.cancer.gov/research/cancer-types/breast-cancer/confluence-project) large-scale genotyping experiment. More detailed fine mapping and functional analysis will be required to elucidate the role of the novel variants identified in BC development for BRCA1 and BRCA2 mutation carriers. Our findings should contribute to the improved performance of BC PRs for absolute risk prediction for BRCA1 and BRCA2 mutation carriers, which will help inform decisions on the best timing for risk-reducing surgery, risk reduction medication, or the start of surveillance.

Methods

Study sample. We used data from two international consortia, BCAC44 and CIMBA46. BCAC included data from 108 studies of BC from 33 countries in North America, Europe and Australia, the majority (88%) of which were case-control studies. The majority of BCAC cases/controls were not tested for BRCA1/2 mutations at the time of enrolment. However, most studies were population-based, hence the proportion of cases and controls that carry pathogenic BRCA1/2 mutations will be small. CIMBA participants were women with pathogenic mutations in BRCA1 or BRCA2. All participants were at least 18 years old. The majority of mutation carriers were recruited through cancer genetics clinics and enrol into national or regional studies. Data were available on 30,500 BCAC BRCA1 mutation carriers and 20,500 BRCA2 mutation carriers from 77 studies in 32 countries. A total of 188,320 BC cases and 161,669 controls were available from both consortia. All studies provided information on disease status, age at diagnosis or at interview. Oestrogen receptor status was available for 72% of BCAC cases and 71% of CIMBA cases. All subjects provided written informed consent and participated in studies with protocols approved by ethics committees at each participating institution.

Sample selection. BCAC cases were women diagnosed with BC. To define disease status in CIMBA participants, women were censored at the first of the following events: age at BC diagnosis, age at ovarian cancer diagnosis, other cancer, bilateral prophylactic mastectomy or age at study recruitment. Subjects censored at a BC diagnosis were considered as cases.

A control-only analysis was carried out to test the independence between the SNPs and the BRCA1 and BRCA2 mutation carrier status. In BCAC, controls were defined as individuals unaffected by BC at study recruitment45. In CIMBA, participants were considered as controls if they were unaffected at recruitment.

Only women of European ancestry were included. To minimise the chance of observing spurious associations due to differences in the distribution of BC cases in the population by tumour characteristics (defined as unselected BC cases), 3,478 BCAC cases from four studies were excluded because they were included in clinical trials based on breast tumour characteristics as HER-2 receptor status (see Supplementary Table 2). Because all the analyses were adjusted for country, to ensure that the number of subjects in each country stratum was large enough, we excluded the CIMBA data from any country for which there were less than ten BC cases with BRCA1 or BRCA2 mutation. Consequently, data from Poland and Russia were excluded from the BRCA2 analyses (Supplementary Table 3). Finally, duplicate subjects between BCAC and CIMBA were excluded from the BCAC data (114 and 80 subjects from the BRCA1 and BRCA2 case-only analyses, respectively; eight subjects from control-only analyses).

A total of 60,212 BCAC cases and 7,257 BRCA1 mutation carrier cases were available for the BRCA1 case-only analyses and 57,725 BCAC cases and 5,097 BRCA2 mutation carrier cases were available for the BRCA2 case-only analyses (Fig. 1). A total of 45,881 BCAC controls and 5,750 BRCA1 mutation carrier controls were available for the BRCA1 control-only analyses and 43,549 BCAC controls and 4,456 BRCA2 mutation carrier controls for the BRCA2 control-only analyses (Fig. 2).

Genotype data. All the study samples were genotyped using the OncoArray Illumina beadchip456. The array includes a backbone of ~260,000 SNPs that provide genome-wide coverage of common variants, together with markers of interest for breast and other cancers identified through GWAS, fine-mapping of known susceptibility regions, and other approaches.

A standard genotype quality control process was followed for both the BCAC and CIMBA samples which have been described in detail elsewhere46,66. Briefly, this process involved excluding SNPs located on chromosomes Y; SNPs with call rates <95%; SNPs with MAF <0.05 and call rate <98%; monomorphic SNPs; and SNPs for which evidence of departure from Hardy-Weinberg equilibrium was observed (P <10−5 based on a country-stratified test).

Genotypes for ~21 million SNPs were imputed for all subjects using the 1000 Genomes Phase III data (released October 2014) as reference panel, as described previously65. Briefly, the number of reference haplotypes used as templates when imputing missing genotypes was fixed to 800 (k_hap = 800). A two-stage imputation approach was used: phasing with SHAPEIT67,68 and imputation with IMPUTE66 using 5 Mb non-overlapping intervals. Genotypes were imputed for all SNPs that were found polymorphic (MAF >0.1%) in either European or Asian populations.

The genome-wide imputation process described above was carried out separately for the BCAC and CIMBA samples. However, this may potentially lead to spurious associations if there are differences in the quality of the imputation (measured using the imputation accuracy r2 metric)7) between the two datasets. To address this, a two-stage approach was employed which involved including only SNPs for which the difference in r2 between the BCAC and CIMBA SNP imputations (Δr2) was minimal relative to their r2 values. SNPs with r2 >0.9 in both BCAC and CIMBA were kept in the analyses only if Δr2 <0.05; SNPs with 0.8 < r2 < 0.9 in both BCAC and CIMBA were kept if Δr2 <0.02 and, SNPs with 0.5 < r2 < 0.8 in both BCAC and CIMBA were kept if Δr2 <0.01. All SNPs with r2 <0.05 in either CIMBA or BCAC were excluded. Only SNPs with a MAF >0.01 in BCAC case were included.

Consequently, 9,072,535 SNPs were included in the BRCA1 analyses (402,336 genotyped and 8,670,199 imputed SNPs) and 9,047,403 SNPs in the BRCA2 analyses (402,397 genotyped and 8,645,006 imputed SNPs).

Case-only and control-only analyses. The comparison of SNP frequency between CIMBA cases and BCAC cases (case-only analyses), or between unaffected CIMBA subjects and BCAC controls (control-only analyses), was performed using logistic regression adjusted for age at BC diagnosis in the case-only analyses and for age at interview for BCAC controls or at censure for CIMBA unaffected subjects in the control-only analyses, as well as for country and principal components (PCs) to account for population structure. Separate analyses were carried out for BRCA1 and BRCA2 mutation carriers. To define the number of PC for inclusion in the models, the principal component analysis was carried out using 35,858 uncorrelated genotyped SNPs on the OncoArray and purpose-written software (http://cgee.medschl.cam.ac.uk/software/pccalc/). The inflation statistic was calculated and converted to an equivalent statistic for a study of 1,000 subjects for each outcome (λm,m) by adjusting for effective study size:

\[ \lambda_{m,m} = \left( \frac{1}{m} \right) + \left( \frac{1}{m} \right) \times \frac{1}{\Delta V} \times \frac{1}{\Delta V} \]

where n and m are the numbers of BCAC and CIMBA subjects respectively. The models were adjusted with the first four PCs (λ1,001 with and without PCs in the model = 1.03 and 1.21, respectively) since additional PCs did not result in further reduction in the inflation of the test statistics.

Strategy for determining significant associations. The analytical process is summarised in Figs. 3 and 4. A fundamental assumption when using a case-only design in this context is that the SNPs and mutation carrier status are independent.41. To confirm independence, SNPs likely to be in linkage disequilibrium (LD) with BRCA1 or BRCA2 mutations, i.e., those located in or within 200 kb of the gene, were excluded. However, LD also exists between variants at long-distance on the same chromosome or even on a different chromosome (interchromosomal LD)41,42,43.
Therefore, control-only analyses were performed to further exclude SNPs associated with mutation carrier status in unaffected women, using a stringent statistical significance threshold of $10^{-8}$. After excluding SNPs in LD or in interchromosomal LD with BRCA1 or BRCA2 mutations, case-only analyses were performed to assess the association between SNPs and BRCA1 or BRCA2 mutation carrier status. We considered two categories of SNPs depending on whether they had been previously found to be associated with BC in published BCAC studies (Fig. 3) and for potential novel SNP modifier (Fig. 4) a stringent significance threshold of $10^{-8}$ was used. Because BRCA1/2-associated tumours are more often ER-negative than those in the general population, we conducted logistic regression analyses using all variants (CCVs) to use in the prediction of the likely target genes. For this purpose, we used a significance threshold of $10^{-4}$ (applying Bonferroni correction to 179 tests) and for potential novel SNP modifier (Fig. 4) a stringent significance threshold of $10^{-4}$ was used.

To confirm that potentially novel associations in the case-only analysis were not driven by differences in the imputation accuracy between the BCAC and BCAC data, each of the regions defined as $ \pm 500 \text{ kb}$ around the associated SNP, were re-imputed using the one-stage imputation applied to BCAC and BCAC data. The more accurate one-stage imputation was carried out, using IMPUTE2 without pre-phasing. Associations with all the SNPs in the re-imputed regions were then re-evaluated using the control-only and case-only analytical approaches described above. Finally, we used a step-wise regression analysis using a significance threshold of $10^{-5}$ in order to determine whether SNPs in the same region are independent and to define the conditionally independent SNPs (top SNPs). Among the 179 established BC susceptibility SNPs, 107 were genotyped and 71 were imputed. As previously, although none of these 71 SNPs were excluded based on their $\Delta^2$, to exclude potentially spurious associations, regions around these 71 SNPs were re-imputed using the one-stage imputation applied to BCAC and BCAC data combined, and before performing the control-only and case-only analyses.

Determination of the magnitude of association. For the potentially novel SNP modifiers the risk ratio of BC applicable to mutation carriers was assumed to be equal to the OR estimate from the case-only analysis (with the hypothesis that their relative risk equals 1 in the general population, given that none of them was found to be associated with BC in BCAC). SNPs, a significant association in the case-only analysis implies that the magnitude of association is different for BRCA1 or BRCA2 mutation carriers than for the general population. Therefore, the risk ratio of BC for mutation carriers was computed as the product of $\text{OR} \times \text{OR}_{\text{BCAC}}$ where OR was obtained from the case-only analysis, and $\text{OR}_{\text{BCAC}}$ was the odds ratio of association obtained from each study. We used the formula $\Delta \text{In}(\text{OR} \times \text{OR}_{\text{BCAC}}) = \text{In}(\text{OR})$, for the SNPs associated with ER-negative BC and from Milne et al. for the SNPs associated with ER-negative BC.

For all associated SNPs in case-only analyses, heterogeneity by country was assessed using likelihood ratio tests that compared models with and without an SNP by country interaction term. When the heterogeneity test was significant at $p < 0.05$, a leave-one-out analysis was performed, by excluding each country in turn to assess the influence of a data from a specific country on the overall association.

Credible causal variants. For each novel region, we defined sets of credible causal variants (CCVs) to use in the prediction of the likely target genes. For this purpose, we defined a first set of CCVs including the top SNP of the region of interest and the SNPs with $p$-values of association within two orders of magnitude of the top SNP association. Then, we sequentially performed logistic regression analyses using all other SNPs that were associated for the top SNP. SNPs associated with CCVs which included the most significant SNP after adjusting for the top SNP and the SNPs with $p$-values within two orders of magnitude of the most significant SNP association. This was repeated (conditioning on the previously found most significant SNPs) to define additional sets of CCVs as long as at least one $p$-value remained $< 10^{-8}$.

eQTL analysis. Data from BC tumours and adjacent normal breast tissue were accessed from The Cancer Genome Atlas (TCGA). Germine SNP genotypes (Affymetrix in 0.61%) and mRNA expression data for individuals of European ancestry were pre-processed and imputed to the 1000 Genomes reference panel (October 2014)33. Tumour tissue copy number was estimated from the Affymetrix 6.0 and called using the GISTIC2 algorithm36. Complete genotype, RNA-seq and copy number data were available for 679 genetically European patients (78 with adjacent normal tissue). Further, RNA-seq data for normal tissue and imputed mRNA expression genotype data were available from 80 females from the GTEx Consortium37. Genes with a median expression level of 0 RPKM across samples were removed, and RPKM values of each gene were log2 transformed. Expression values of samples were quantile normalized. Genetic variants were evaluated for association with the expression of gene expression under the lead variant within each individual using a computational pipeline, INeurolog expression QuAntitative trait and In Silico prediction of GWAS Targets (INQUIST33). Briefly, genes were considered as potential targets of candidate causal variants through effects on: (1) distal gene regulation, (2) proximal regulation, or (3) a gene's coding sequence. We intersected CCV positions with multiple sources of genomic information chromatin interaction analysis by paired-end tag sequencing (ChIA-PET97) in MCF7 cells and genome-wide chromosome conformation capture (Hi-C) in HMECs. We used breast cell line computational enhancer–promoter correlators (PreSTIGE80, IM-PET81, FAN- TOMS82) breast cell super-enhancer83, breast tissue-specific expression variants (eQTL) from multiple independent studies (TCGA (normal breast and breast tumour) and GTEx breast—see eQTL methods), transcription factor and histone modification chromatin immunoprecipitation followed by sequencing (ChIP-seq) from the ENCODE and Roadmap Epigenomics Projects together with the genomic features found to be significantly enriched for all known breast cancer CCVs84, computational prediction methods, or eQTL associations. All CCVs were considered as potentially involved in distal interaction. Intersection of a putative distal enhancer with genomic features found to be signficantly enriched84 were further upweighted. Multiple independent interactions were awarded an additional point. CCVs in gene proximal regulatory regions were intersected with histone ChIP-seq data and enhancer regions were categorized based on adh and if scores. Nonsense and missense changes were assessed with the REVEL ensemble algorithm, with CCVs displaying REVEL scores $> 0.5$ deemed deleterious.

Each target gene prediction category (distal, promoter or coding) was scored according to different criteria. Genes predicted to be distally regulated targets of CCVs were awarded points based on physical links (for example ChIA-PET), computational prediction methods, or eQTL associations. All CCVs were considered as potentially involved in distal association. Intersection of a putative distal enhancer with genomic features found to be significantly enriched were further upweighted. Multiple independent interactions were awarded an additional point. CCVs in gene proximal regulatory regions were intersected with histone ChIP-seq data and enhancer regions were categorized based on adh and if scores. Nonsense and missense changes were assessed with the REVEL ensemble algorithm, with CCVs displaying REVEL scores $> 0.5$ deemed deleterious.

Data availability. All data and resources used in the study, including the INQUIST pipeline, the ENCODE and Roadmap Epigenomics Projects, and the REVEL ensemble algorithm are available on the INQUIST website (http://inquisit.org). The complete dataset is provided in the Supplementary Table 2 and is available on request from the corresponding author. The data were also made available to the Data Access Coordinating Committees (DACCs) of the BCAC and BRCA1/2 Consortiums (BRCA2 Consortium 47, 2014 Genome wide association study). Requests for the complete dataset can be made to the corresponding author or the Access Committee of the study (BRCA and BCAC).
HERCS, HEBON, IHCC, INHERIT, I01HBRCs, MCGILL, NRG, ONCOLOGY, OUI and UKGFOCR (see Supplementary Table 1—for a list of all CIMBA studies). Case-control summary results from CIMBA and BCAC consortia are publicly available and can be downloaded at http://cimba.cge.medschl.cam.ac.uk/oncoarray-complete-summary-results/ and at http://bcac.cge.medschl.cam.ac.uk/bcadata/oncoarray/oncoarray-and-combined-summary-result/; gwas-summary-associations-breast-cancer-risk-2020/. The top 10 000 SNPs from the current BCAC-CIMBA case-only study can be found at http://cimba.cge.medschl.cam.ac.uk/projects/BCAC-CIMBA_Case_only_analysis. The remaining data are available within the Article, Supplementary Information or available from the authors upon request. Source data are provided with this paper.

Received: 28 October 2019; Accepted: 19 November 2020; Published online: 17 February 2021

References

1. Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods, and major patterns in GLOBOCAN 2012. Int. J. Cancer 136, E359–E386 (2015).
2. Pharoah, P. D. P., Day, N. E., Duffy, S., Easton, D. F. & Ponder, B. A. J. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int. J. Cancer 71, 800–809 (1997).
3. Nelson, H. D. et al. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: a systematic review to update the U.S. Preventive Services Task Force recommendation. Ann. Intern. Med. 160, 255–266 (2014).
4. Kuchenbaecker, K. B. et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA 317, 2402–2416 (2017).
5. Antoniou, A. C. et al. Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers. Hum. Mol. Genet. 18, 4442–4456 (2009).
6. Antoniou, A. C. et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nat. Genet. 42, 885–892 (2010).
7. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45, 353–361 (2013).
8. Thomas, G. et al. A multi-stage genome-wide association in breast cancer identifies two novel risk alleles at 1p11.2 and 1q42.4 (RAD51L1). Nat. Genet. 41, 579–584 (2009).
9. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).
10. Garcia-Closas, M. et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat. Genet. 45, 392–398e2 (2013).
11. Couch, F. J. et al. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. Nat. Commun. 7, 11375 (2016).
12. Lin, W.-Y. et al. Identification and characterization of novel associations in the CASP8/ALS2CR12 region on chromosome 2 with breast cancer risk. Mol. Genet. Genom. 51, 285–298 (2015).
13. Milne, R. L. et al. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. Hum. Mol. Genet. 23, 6096–6111 (2014).
14. Haiman, C. A. et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat. Genet. 43, 1210–1214 (2011).
15. 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, 371–384e2 (2013).
16. Ghousaini, M. M. et al. Evidence that the Sp12 variant rs10941679 confers susceptibility to estrogen-receptor-positive breast cancer through FGFI0 and MRPS30 regulation. Am. J. Hum. Genet. 99, 903–911 (2016).
17. Glubb, D. M. et al. Fine-scale mapping of the 1q21.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. Nat. Genet. 21, 5373–5384 (2012).
18. Dienstmann, R. et al. Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nat. Genet. 48, 374–386 (2016).
19. Sawyer, E. et al. Genetic predisposition to in situ and invasive lobular carcinoma of the breast. PLoS Genet. 10, e1004285 (2014).
20. Easton, D. F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087–1093 (2007).
21. Dunning, A. M. et al. Breast cancer risk variants at 6q14.3 and 12p13.13 are associated with breast cancer risk in BRCA1 and/or BRCA2 mutation carriers. Breast Cancer Res. 14, R33 (2012).
22. Chenevix-Trench, G. et al. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). Breast Cancer Res. 9, 104 (2007).
23. Sawyer, E. et al. Genetic predisposition to in situ and invasive lobular carcinoma of the breast. PLoS Genet. 10, e1004285 (2014).

52. Ottman, R. Gene–environment interaction: definitions and study designs. Prev. Med. 25, 76–102 (1996).

53. Andrieu, N. & Goldstein, A. M. Epidemiologic and genetic approaches in the study of gene-environment interaction: an overview of available methods. Epidemiol. Rev. 20, 137–147 (1998).

54. Fachl, L. et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nature genetics, 52, 56–73 (2020).

55. Whittemore, A. S. Assessing environmental modifiers of disease risk associated with rare mutations. Hum. Hered. 63, 134–143 (2007).

56. CIMBA - Consortium of Investigators of BRCA1/2. — http://cimba.cge.medschl.cam.ac.uk/ (2017).

57. Turner, A. et al. MADD knockdown enhances doxorubicin and TRAIL-induced apoptosis in breast cancer cells. PLoS ONE 8, e68817 (2013).

58. Zhong, Q., Yang, A., Hu, D. & Wang, Y. Molecular mechanisms of breast cancer metastasis by gene expression profile analysis. Mol. Med. Rep. 16, 4671–4677 (2017).

59. Sharma, D. K., Bressler, K., Patel, H., Balasingam, N. & Thakor, N. Role of Eukaryotic initiation factors during cellular stress and cancer progression. J. Nucl. Acids. 2016, 8235121 (2016).

60. HANNA, S. & SLETON-Lee is a tumor suppressor in breast cancer that regulates cell motility and invasion. Int. J. Oncol. 44, 1499–1511 (2014).

61. Piepersch, W. W., Weinberg, C. R. & Taylor, J. A. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. Stat. Med. 13, 153–162 (1994).

62. Hohenlohe, P. A., Bassham, S., Currey, M. & Cresko, W. A. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos. Trans. R. Soc. B Biol. Sci. 367, 395–408 (2012).

63. Umbach, D. M. & Weinberg, C. R. Designing and analysing case-control studies to exploit independence of genotype and exposure. Stat. Med. 16, 1731–1743 (1997).

64. Spurdle, A. B. et al. Refined histopathological predictors of BRCA1 and BRCA2mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. Breast Cancer Res. 16, 3419 (2014).

65. McLaren, W. et al. The ensembl variant effect predictor. Genome Biol. 17, 122 (2016).

66. nature.com/naturecommunications

Acknowledgements

BCAC acknowledgments. We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. ARCS thank Maggie Angelakos, Judi Maskell, Gillian Dite. ARCS thanks the Blood bank Sanquin, The Netherlands. AACTB Investigators: Christine Clarys, Yvonne Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirmala Pathanmanathan, Peter Simpson, J. Dinny Graham, Mythily Sachitananthan. Samples are made available to researchers on a non-exclusive basis. BCS thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. BCEES thanks Allyson Thomson, Christobel Sauls, Terry Slevin, BreastScreen Western Australia, Elizabeth Wyche, Rachel Lloyd. The BCNIS study would not have been possible without the contributions of Dr. K. Landman, Dr. N. Gronich, Dr. A. Flugelman, Dr. K. Sainsbury, Dr. E. Liani, Dr. I. Cohen, Dr. S. kalet, Dr. V. friedman, Dr. O. barnet of the N ICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. The BREOGAN study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelas, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Dominguez, MARCS thank Ana Martinez, Sara Minaya Ponce, Carmen Redondo Fer- nández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BCROGÁN), José Antínuez, Maximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS; Joaquín González-Campos and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain. BSUCH thanks Peter Bugert, Medical Faculty Mannheim. BCBS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. CCGB thanks Stylian Apostolaki, Anna Margioliński, Georgios Nintos, Maria Perraki, Georgia Solastrouorou, Giorgos Kontossilis, Konstantinos Pompeadis. GPGS thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. CNIO-BCS thanks Guillermo Pita, Chao Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotypting CEGENE Unit (CNIO). The CTS Steering Committee includes Leslie Bernstam, Michael Kastan, James Lacey, Sophia Wang, Huiyan Ma, and Jessica Clague DeHart at the Beckman Research Institute of City of Hope, Dennis Deapan, Rich Pinder, and Eunjung Lee at the University of Southern California, Pam Horn-Ross, Peggy Reynolds, Christina Clarke Dur and David Nelson at the Cancer Prevention Institute of California, Hoda Anton- Culver, Argyrios Zougas, and Hannah Park at the University of California Irvine, and Fred Schumacher at Case Western University. DIETCOMFLPY thanks the patients, nurses and clinical staff involved in the study. The DietComFLpy study was funded by the charity Against Breast Cancer (Registered Charity Number 1121258) and the NCRN. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks Harriett Ziegler, Sonja Wolf, Volker Her- manmann, Susanne Stegmaier, Katja Rutteger, BCCA (HRCC) thanks Stefan Engert, Heide Hellebrand, Sandra Krober and LIFE - Leipzig Research Centre for Civilisation Diseases (Markus Loefker, Joachim Thiery, Matthias Nüchter, Ronny Baber). The GENICA Network: Dr. Margarete Fischer-Bosch-Institut of Clinical Pharmacology, Stuttgart, and University of Tubingen, Germany [HB, Wing-Yee Lo], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) Partner Site Tübingen [HB], gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzeller-enzstrategie des Bundes und der Länder – EXC 2180 – 390900677 [HB], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [YDK, Christian Baesch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [Thomas Bruning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. HBRCs thanks Michael Bremer. HERCS thanks Kristian Aaltonen, Ilya Erikkila. HABRA thanks Shamil Gantsetseg and SASCAB thank the Swedish Medical Research Council. KRCP thanks Eija Myölänen, Helena Kemiläinen. kConFab/AOCS wish to thank Heather Thorné, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow-Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, the Victorian Breast Health Research Foundation, and the National Breast Cancer Foundation). We also thank the many families who contribute to kConFab Resource. LMBRC thanks Gilian Peutenman, Thomas Van Brussel, EvyVanderheynden and Kathleen Corhouts. MARIE thanks Petra Seiboldt, Dieter Flesch-Jany, Judith Heinz, Nadia Obi, Alina Vrieling.

84. Dixon, J. et al. Integrative detection and analysis of structural variation in cancer genomes. Nat. Genet. 50, 1388–1389 (2018).

85. McLaren, W. et al. The ensemble variant effect predictor. Genome Biol. 17, 122 (2016).

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-20496-3 | www.nature.com/naturecommunications
Lisa Golmard, Claude Houdayer, Marine Le Mentec, Virginie Moncoutier, Antoine de Buecher, Sandrine Caputo, Anaïs Dupré, Emmanuelle Fourme, Marion Gauthier-Villars, resource centre was transferred to Paris in December 2015 (Noura Mebirouk, Fabienne coordinated GEMO until she sadly passed away on the 30th June 2014. The team in Lyon to Olga M. Sinilnikova, who with Dominique Stoppa-Lyonnet initiated and Cancer Risk in BRCA1/2 Mutation Carriers (GEMO) study is a study from the National Thiery, Matthias Nüchter, Ronny Baber); We thank all participants, clinicians, family participants as required. We would like to thank the participants and staff of the NHS and NHSD for their valuable contributions as well as the following state cancer registries for their help: A.L, A.Z, A.R, A.C, C.O, C.T, D.E, F.L, G.A, I.D, H.I., I.N, I. A., K.Y, I.A, M.E, M.D, M.A, M.I, N.E, N.H, N.J, N.Y, N.C, N.D, O.H, O.K, O.R, P.A., R.L, S.C, T.N, T.X, V.A, W.A, and W.Y. The authors assume full responsibility for analyses and interpretation of these data. OFBCR thanks Teresa Selandar, Yanaya Weerasooriya. ORIGO thanks E. KroI Warmerdam, and J. Blom for patient accrual, administrative and clinical support, and managing the data. JHCC thanks Xavier Brion, Brinton, Mark Sherman, Neolina Szieszzenia-Dubrowska, Beata Pepelnos, Witold Zatoraki, Pei Chao, Michael Stagner. The ethical approval for the POSI study is MREC /006/69, UKCRN ID: 1137. We thank our Experimental Cancer Medicine Centre (ECMC) supported Faculty of Medicine Tissue Bank and the Faculty of Medicine DNA Banking resource. RBCC thanks Jannet Blom, Saskia Felders, Annette Hremskerk and the Erasmus Centre of Molecular Diagnostics. We thank the Department of Pathology of the Erasmus MC, and Linda Steele for their work in participant enrollment and biospecimen and data management; Bent Ejelftten and Anne-Marie Gerdes for the recruitment and genetic counseling of participants; Alicia Barroso, Rosario Alonso and Guillermo Pita; all the individuals and the researchers who took part in CONSIT TEAM (Consortio Italiano Tumori Ereditari Alle Mammella), in particular: Bernard Peissel, Dario Zimbalatti, Daniela Zaffaroni, Alessandra Viel, Giuseppe Giannini Liliana Varesco, Viviana Gis... It has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre; Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, and Pedro Pinto; members of the Center of Molecular Diagnosis, Oncogenetics Dumont for sample management and skillful assistance; Ana Peixoto, Catarina Santos and Pedro Pinto; members of the Center of Molecular Diagnosis, Oncogenetics Dumont for sample management and skillful assistance; Ana Peixoto, Catarina Santos... We thank the Department of Molecular Oncology Research Center of Barretos Cancer Hospital; Heather Thorne, Eveline Niedermayer, all the KConFab research nurses and staff, and the staff of the Family Cancer Clinics, and the Clinical Follow-Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA) for their contributions to this research and the many families who contribute to the KConFab), the investigators of: Guy-Bonneau–Ferrand, France: Yves-François, CHU Arnaud-de-Villeneuve, Montpellier, France: Isabelle Couper, Pascal Pujol, Centre Oscar Lambret, Lille, France: Claude Adenis, Aurélie Dumont, France Révillon. Groupe Cancer Familial SPECTIVE project supported by the Government of Canada through Genome Canada – France: Valérie Bonadona, Centre François Baclesse, Caen, France: Pascale Bellet, Laurent Castro, Dominique Vuar. Institut Paoli Calmettes, Marseille, France: Violaine Bourbon, Catherine Nozé, Tetsuro Sato, and the population recruited in the LMH. The GEMO study followed the Declaration of Helsinki and Good Clinical Practice Guidelines. The National Cancer Genetics Network; Leigha Senter, Kevin Sweet, Caroline Craven, Julia Cooper, and Michelle O’Conor; HVH: acknowledgments to the Cellex Foundation for providing research facilities and equipment. Dr Juliette Coignard was supported by a fellowship of INCA Institut National du Cancer No2015-181, a Ligue Nationale contre le Cancer IP/SC-15229 and Olga Sinilnikova fellowship of INCa Institut National du Cancer N°2015-181, la Ligue Nationale contre le Cancer IP/SC-15229 and Olga Sinilnikova’s fellowship (2016). BCAC Funding, BCAC is funded by Cancer Research UK [C128/18563]. Ligue Nationale contre le Cancer IP/SC-15229 and Olga Sinilnikova’s fellowship (2016). BCAC Funding, BCAC is funded by Cancer Research UK [C128/18563], C128/18(A1018), the European Union’s Horizon 2020 Research and Innovation Programme (grant numbers 639435 and 633784 for BRIDGES and B-CAST respectively), and by the European Community’s Seventh Framework Programme under grant agreement number 221375 (grant number H2020-2014-2015). The Canadian Breast Cancer Family Cancer Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating institutions. The authors declare no competing interests. Funding Sources: commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Program...
Council of Australia, the New South Wales Cancer Council, the Victorian Health Pro-
motion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J. L.H. is the Pacific Director for the National Cancer Institute, National Institutes of Health, for the Breast Cancer Family Registry (BCFR), M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society [grants NKI 2007-3838; 2009 4363]. The Aus-
tralian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, the Cancer Institute NSW and the National Breast Cancer Foundation. The German Breast Cancer Society (DBK) and the Finnish Breast Cancer Research Foundation were supported by the University Hospital of Erlangen. The BCRS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BCESS was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foun-
dation. The BCSC was supported by the Danish Cancer Society. The HER-2 Breast Cancer Study was funded by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. The BiEast Oncology Gailiatic Network (BriegsG mik) was funded by the Acces al Escontria de Salud del Instituto de Salud Carlos III FIS PI12/01255/ 
Cofaniciado FEDIER: Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI13/01136; Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestión Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Pro-
grama de Xestión de Centros de Experiencia Aplicada, Plan 2009-2013 GEXI/10; 11D INE BID858/13. The Breastcancer Support Study was supported by the National Institute of Health and Medical Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Kanker, De Ruijter Foundation, The Northern Cancer Database. The MEC was support by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study is funded by support from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Krapmand Foundation, Gunnar Nilsson. The MMHS study was supported by NIH grants CA79396, CA128931, CA162001, CA140286 and CA177150. MRCC was supported by grants from the Breast Cancer Research Foundation and Rundkamp's Cancer Research Fund (Rundkamp's Cancer Research Fund). The work of MLTGBECs was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the CIHR Team in Familial Risks of Breast Cancer program – grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade – grant # PSR-SIRI-701. The NHBS was supported by NIH grant R01CA103874. Biological sample preparation was conducted the Survey and Biomarker Research (SBR), which is supported by R01 CA81898. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grant U10 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The Carolina Breast Cancer Study was funded by Komen Foundation, the National Cancer Institute (PS0 CA058223, US5 CA156733, U10 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, U101 CA186107, and U19 CA48065. The NHBS was supported by NIH grants P01 CA103874 and U19 CA48065. The OFBCR was supported by grant U10 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The Carolina Breast Cancer Study was funded by Komen Foundation, the National Cancer Institute (PS0 CA058223, US5 CA156733, U10 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, U101 CA186107, and U19 CA48065. The NHBS was supported by NIH grants P01 CA103874 and U19 CA48065. The OFBCR was supported by grant U10 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR.
the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NIH in the East of England through the Clinical Academic Reserve. The Sixth Cancer Research UK Hardcover Campaign, the National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The Two Sister Study (2SISTER) was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES012245), and, also by a grant from Susan G. Komen for the Cure, grant FAS7030856. SKDRC is supported by the DKFZ. The Fisher Cancer Foundation and the Swedish Research Council (VR 2017-00644) for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The SZBCS and IHCC were supported by Grant PBZ_KBN_122/P05/2004 and the program of the Minister of Science and Higher Education under the name Regional Initiative of Excellence in 2019–2022 project number 002/RID/2018/19 amount of 2,100,000 PLN. The TNBCG is supported by the Spanish Society of Research Excellence (SORE) in Breast Cancer (CA116201), a grant from the Breast Cancer Research Foundation, a generous gift from the David F. and Margaret T. Grohe Family Foundation. The UCIBCS component of this research was supported by the NIH [CA58860, CA92044] and the Lon V Smith Foundation [LVS93492]. The UKBGS was funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. The UKBGS study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Bio-medical Research Centre. CIMBA Funding. CIMBA: The CIMBA data management and data analysis were supported by Cancer Research – UK grants C11229/A20861, C12292/ A11174. GCT and ABS are NHMRC Research Fellows. iCOGS: The European Community Sixth Framework Programme (2003-2006) (LIFTH- F2-2009-223175) (COGS), Cancer Research UK (C1287/A11018, C1287/A10710, C12292/A11174, C12821/A10214, C5047/A8384, C5047/A10057, C5047/A10692, C18197/ A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (U19 CA148537, U19 CA148065 and U19 CA148112 – the GAME-ON initiative), the Department of Defence (W181X1-10-1-0341), the Canadian Institutes of Health Research (CIHR), the CHR (Team Health Outcomes) of the National Cancer Institute of Canada, the Ministry of Economic Development, Innovation and Export Trade (PSR-SIIRI-701), Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The PERSPECTIVE project was supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministry of Economy, Science and Innovation through Genome Quebec, and The Quebec Breast Cancer Foundation. BCRF, UM1 CA146820 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCRF.

BIDMC: Breast Cancer Research Foundation. CNIO: Spanish Ministry of Health P16/11-0440 supported by FEDER funds, the Spanish Ministry of Economy and Competitiveness (MINECO) SAF2014-57680-R and the Spanish Research Network on Rare diseases (CIBERER). COH-CCGRN: Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under grant R25CA112486, and RC4CA153828 (PI: J. Wetzelt) from the National Cancer Institute (NCI). The Stanford University Cancer Center Program of Research Excellence (SPORE) in Breast Cancer is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. CONSIT TEAM: Funds from Italian citizens who allocated the 5 x 1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects ‘5 x 1000’) to S. Manoukian. Associazione Italiana Ricerca sul Cancro (AIRC, IG2015 no.16732) to P. Petroni. DEMOKRITOS: European Union (European Social Fund – ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program of the General Secretariat for Research & Technology: SYN11_10_19 NBCA. Investing in knowledge society through the European Social Fund. DFKZ: German Cancer Research Center. EMBRACE: Cancer Research UK Grants C1287/ A10118 and C1287/A11990. D. Gareth Evans and Fiona Lalloo are supported by an NIHR grant to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. FCCC: A.K.G. was in receipt of a Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN) and the Canadian Institutes of Health Research (grants 17-00-00171, 18-515-2452 and 19-515-2501). NNPIO: the Russian Foundation for Basic Research (grants 17-00-00171, 18-515-45012 and 19-515-2501). NRG Oncology: U10 CA180868, NRG SDMC grant U10 CA180822, NRG Administrative Office and the NRG Tissue Bank (CA 27469), the NRG Statistical and Data Center (CA 37517) and the Intramural Research Program, NCI. OSUCCG: Ohio State University Comprehensive Cancer Center. PBCS: Italian Association of Cancer Research [AIRC] [IG 2013 N.14477 and Tuscany Institute for Tumours (ITT) grant 2014-2015-2016. SMC: the Israeli Cancer Association. Sweden-BREA: the Swedish Cancer Society. UCHICAGO: NCIC Specialized Program of Research Excellence (SORE) in Breast Cancer (CA116201) and a grant from the Breast Cancer Research Foundation. McGill: General Hospital Weekend to End Breast Cancer, Quebec Ministry of Economic Development, Innovation and Export Trade. Marc Tischkowitz is supported by the funded by the European Union Seventh Framework Program (2007-2013)/Eur- open Research Council (Grant no. 310018). MSKCC: the Breast Cancer Research Foundation, the Robert and Kate Niehaus Clinical Cancer Genetics Initiative, the Andrew Saban Research Fund and a Cancer Center Support Grant/Core Grant (P30 CA08748). NCI: the Intramural Research Program of the US National Institutes of Health, NIH, and by support services contracts N02-CP-11019-50, N02-CP-2103-63 and N02- CP-65504 with Westat, Inc, Rockville, MD. NNPIO: the Russian Foundation for Basic Research (17-00-00171, 18-515-45012 and 19-515-2501). NRG Oncology: U10 CA180868, NRG SDMC grant U10 CA180822, NRG Administrative Office and the NRG Tissue Bank (CA 27469), the NRG Statistical and Data Center (CA 37517) and the Intramural Research Program, NCI. OSUCCG: Ohio State University Comprehensive Cancer Center. PBCS: Italian Association of Cancer Research (AIRC) [IG 2013 N.14477 and Tuscany Institute for Tumours (ITT) grant 2014-2015-2016. SMC: the Israeli Cancer Association. Sweden-BREA: the Swedish Cancer Society. UCHICAGO: NCIC Specialized Program of Research Excellence (SORE) in Breast Cancer (CA125183), R01 CA42996, 1U01CA161632 and by the Ralph and Marion Falk Medical Research Trust, the National Cancer Institute of Biomedical and Neurological Women’s Cancer Research Alliance and the Breast Cancer research Foundation. UCSF: UCSF Cancer Risk Program and Helen Diller Family Comprehensive Cancer Center. UPPEN: Breast Cancer Research Foundation; Susan G. Komen Foundation for the cure, Basier Research Center for BRCA. UPITT/ MWH: Hackers for Hope Pittsburgh. VFTCG: Victorian Cancer Agency, Cancer Aus- tralia, National Breast Cancer Foundation. WGF: Dr Karlan is funded by the American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN) and the National Center for Advancing Translational Sciences (NCATS), Grant ULTR000124. HVH: Supported by the Carlos III National Health Institute funded by FEDER funds – a way to build Europe – P16/11363. MT Parsons is supported by a grant from Newfoundland University. Kelly-Anne Phillips is an Australian National Breast Cancer Foundation Fellow.

Author contributions

A.C.A., D.F.E. and N.A. conceived the study design. J.C., N.A. and A.C.A. drafted the manuscript, while the complete writing group consisted of J.C., N.A., A.C.A., G.C., T. and D.F.E. performed the statistical analyses and J.B. and T.A.O. the INQUISIT predictions. M.L. and J.P.T. contributed to bioinformatics analyses. D.R.B. contributed to statistics analyses. J.D., L.M., G.L. and M.K.B. performed the data management. M.A.A., S.A., T.A., K.A., I.A., H.A.-C., V.A., N. Arnold, J.K., B.A.A., A.A., J.A., C.B., H.B., M. B., L.B., K.B., C. Blomqvist, D.B., S.E.B., B.B., A.B., H.B., Brenner, B.Wurkel, S.B. and T.C. performed the statistical analyses. T.C., M.D.C., K.B., P.D.B., O.D., Y.C.D., S.M.D., T.D., I.D-.S-.S., A.M.D., M.D., D.M.E., A.H.E., C.E., M.E., D.G.E., P.A.F., H.F., E.F., L.F., D.F., M.G.-D., S.M.G., J.G., V.-G.
Juliette Coignard1,2,3,4,5,6, Michael Lush4, Jonathan Beesley7, Tracy A. O'Mara6,7, Joe Dennis6,4, Jonathan P. Tyrer6,8, Daniel R. Barnes4,4, Lesley McGuffog4, Goska Leslie6,4, Manjeet K. Bolla4, Muriel A. Adank9, Simona Agata10, Thomas Ahearn11, Kristiina Aittomäki12, Irene L. Andrulis13,14, Hoda Anton-Culver15, Volker Arndt16, Norbert Arnold17,18, Kristan J. Aronson19, Banu K. Arun20, Annelies Augustinsson21, Jacopo Azzollini22, Daniel Barrowdale6,4, Caroline Baynes8, Heiko Becher23, Marina Bermisheva24, Leslie Bernstein25, Katarzyna Bialkowska26, Carl Blomqvist27,28, Stig E. Bojesen29,30,31, Bernardo Bonanni32, Ake Borg33, Hiltrud Brauch34,35,36, Hermann Brenner37,38, Barbara Burwinkel39,40, Saundra S. Buys41, Trinidad Caldés42, Maria A. Caligo43, Daniele Campa44,45, Brian D. Carter46, Jose E. Castelao47, Jenny Chang-Claude45,48, Stephen J. Chanock49,11, Wendy K. Chung49,50, Kathleen B. M. Claes51, Christine L. Clarke51, GEMO Study Collaborators*, EMBRACE Collaborators*, J. Margriet Collée52, Don M. Conroy4, Kamila Czene53, Mary B. Daly54, Peter Devilee55,56, Orland Diez57,58, Yuan Chun Ding59, Susan M. Domchek59, Thilo Dörk60, Isabel dos-Santos-Silva61, Alison M. Dunning6,8, Miriam Dwek62, Diana M. Eccles63, A. Heather Eliassen64,65, Christoph Engel66, Mikael Eriksson53, D. Gareth Evans67,68, Peter A. Fasching69,70, Henrik Flyger71, Florentia Fostira72, Eitan Friedman73,74, Lin Fritschi75,76,77, Debra Frost4,4, Manuela Gago-Dominguez68,69, Mary S. Gapstur53, Judy Garber78, Andrew K. Godwin86, Mark S. Goldberg87,88, David E. Goldgar89, Anna Gonzalez-Neira90, Mark H. Greene91, Pascal Guénel92, Lothar Haebeler93, Eric Hahnen94,95, Christopher A. Haiman96, Niclas Håkansson97, Per Håll98, Ute Hamann99, Patricia A. Harrington8, Steven N. Hart100, Wei He101, Frans B. L. Hogervorst92, Antoinette Hollestelle101, John L. Hopper84, Darling J. Horcasitas102, Peter J. Hulick103,104, David J. Hunter105,106, Evgeny N. Imyanitov107, KConFab Investigators*, HEBON Investigators*, ABCTB Investigators*, GEMO Study Collaborators, EMBRACE Collaborators provided DNA samples and/or phenotypic data. All authors read and approved the final manuscript. The funders had no role in the design of the study, the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript for publication.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-020-20496-3.

Correspondence and requests for materials should be addressed to N.A. or A.C.A.
Renske Keeman, Elza Khnusutdinova, Johanna I. Kiiski, Yon-Dschun Ko, Veli-Matti Kosma, Peter Kraft, Allison W. Kurian, Yon-Dschun Ko, Adria Lopez-Fernández, Jennifer T. Loud, Craig Luccarini, Arto Manneuress, Siranoush Manoukian, Sara Margolin, John W. M. Martens, Noura Mebirok, Alfons Meindl, Austin Miller, Roger L. Milne, Marco Montagna, Katherine L. Nathanson, Susan L. Neuhausen, Heli Nevanlinna, Finn C. Nielsen, Katie M. O'Brien, Olufumilayo I. Olopade, Janet E. Olson, Håkan Olsson, Ana Osorio, Laura Ottini, Tjoung-Won Park-Sim, Michael T. Parsons, Inge Sokilde Pedersen, Beth Peshkin, Paolo Peterlongo, Julian Peto, Paul D. P. Pharoah, Kelly-Anne Phillips, Eric C. Polley, Bruce Poppe, Nadege Presneau, Miquel Angel Pujana, Kevin Punie, Paolo Radice, Johanna Rantalä, Muhammad U. Rashid, Gud Rennert, Hedy S. Rennert, Mark Robson, Atocha Romero, Maria Rossing, Emmanouil Saloustrous, Dale P. Sandler, Regina Santella, Maren T. Scheune, Marjanka K. Schmidt, Gunnar Schmidt, Christopher Scott, Priyanka Sharma, Penny Soucy, Melissa C. Southey, John J. Spinelli, Zoe Steinsnyder, Jennifer Stone, Dominique Stoppa-Lyonnet, Anthony Swedlow, Rulla M. Tamimi, William J. Tapper, Jack A. Taylor, Mary Beth Terry, Alex Teule, Darcy L. Thull, Marc Tischkowitz, Amanda E. Tolland, Diana Torres, Alison H. Trainer, Therése Truong, Nadine Tung, Celine M. Vachon, Ana Vega, Joseph Vija, Qin Wang, Barbara Wappenschmidt, Clarice R. Weinberg, Jeffrey N. Weitzel, Camilla Wendt, Alicja Wolk, Siddhartha Yadav, Xiaohong R. Yang, Drakoluis Yannoukakos, Wei Zheng, Argyrios Zogas, Kristin K. Zorn, Sue K. Park, Barbara Wang, Madis Thomassen, Kenneth Offit, Rita K. Schmutzler, Fergus J. Couch, Jacques Simard, Georgia Chenex-v-Trench, Douglas F. Easton, Nadine Andrieu, Fabienne Lesueur, Håkan Olsson, Ana Osorio, Laura Ottini, Tjoung-Won Park-Sim, Michael T. Parsons, Inge Sokilde Pedersen, Beth Peshkin, Paolo Peterlongo, Julian Peto, Paul D. P. Pharoah, Kelly-Anne Phillips, Eric C. Polley, Bruce Poppe, Nadege Presneau, Miquel Angel Pujana, Kevin Punie, Paolo Radice, Johanna Rantalä, Muhammad U. Rashid, Gud Rennert, Hedy S. Rennert, Mark Robson, Atocha Romero, Maria Rossing, Emmanouil Saloustrous, Dale P. Sandler, Regina Santella, Maren T. Scheune, Marjanka K. Schmidt, Gunnar Schmidt, Christopher Scott, Priyanka Sharma, Penny Soucy, Melissa C. Southey, John J. Spinelli, Zoe Steinsnyder, Jennifer Stone, Dominique Stoppa-Lyonnet, Anthony Swedlow, Rulla M. Tamimi, William J. Tapper, Jack A. Taylor, Mary Beth Terry, Alex Teule, Darcy L. Thull, Marc Tischkowitz, Amanda E. Tolland, Diana Torres, Alison H. Trainer, Therése Truong, Nadine Tung, Celine M. Vachon, Ana Vega, Joseph Vija, Qin Wang, Barbara Wappenschmidt, Clarice R. Weinberg, Jeffrey N. Weitzel, Camilla Wendt, Alicja Wolk, Siddhartha Yadav, Xiaohong R. Yang, Drakoluis Yannoukakos, Wei Zheng, Argyrios Zogas, Kristin K. Zorn, Sue K. Park, Barbara Wang, Madis Thomassen, Kenneth Offit, Rita K. Schmutzler, Fergus J. Couch, Jacques Simard, Georgia Chenex-v-Trench, Douglas F. Easton.
Department of Human Genetics University of Leuven, Leuven, Belgium. 127Epidemiology Program University of Hawaii Cancer Center, Honolulu, HI, USA. 128High Risk and Cancer Prevention Group Vall d’Hebron Institute of Oncology, Barcelona, Spain. 129Department of Clinical Science and Education, Södersjukhuset Karolinska Institutet, Stockholm, Sweden. 130Department of Gynecology and Obstetrics University of Munich, Campus Grosshadern, Munich, Germany. 131INRC Oncology, Statistics and Data Management Center Roswell Park Cancer Institute, Buffalo, NY, USA. 132Center for Genomic Medicine Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. 133Epidemiology Branch National Institute of Environmental Health Sciences, NIH Research Triangle Park, Durham, NC, USA. 134Center for Clinical Cancer Genetics The University of Chicago, Chicago, IL, USA. 135Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain. 136Department of Molecular Medicine University La Sapientia, Rome, Italy. 137Molecular Diagnostics Aalborg University Hospital, Aalborg, Denmark. 138Clinical Cancer Research Center Aalborg University Hospital, Aalborg, Denmark. 139Department of Clinical Medicine Aalborg University, Aalborg, Denmark. 140Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA. 141Genome Diagnostics Program IFOM - the FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Milan, Italy. 142Translational Research Laboratory IDIBELL (Bellvitge Biomedical Research Institute), Catalan Institute of Oncology, CIBERONC, Barcelona, Spain. 143Leuven Oncology Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium. 144Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy. 145Clinical Genetics Karolinska Institutet, Stockholm, Sweden. 146Department of Basic Sciences Sha'akun Hamid Memorial Cancer Hospital and Research Centre (SKMCH & RC), Lahore, Pakistan. 147Clalit National Cancer Control Center Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel. 148Clinical Genetics Service, Department of Medicine Memorial Sloan-Kettering Cancer Center, New York, NY, USA. 149Medical Oncology Department Hospital Universitario Puerta de Hierro, Madrid, Spain. 150Department of Oncology University Hospital of Larissa, Larissa, Greece. 151Department of Epidemiology, Mailman School of Public Health Columbia University, New York, NY, USA. 152Cancer Genetics and Prevention Program University of California San Francisco, San Francisco, CA, USA. 153Division of Psychosocial Research and Epidemiology The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands. 154Institute of Human Genetics Hannover Medical School, Hannover, Germany. 155Department of Internal Medicine, Division of Medical Oncology University of Kansas Medical Center, Westwood, KS, USA. 156Genomics Center, Centre Hospitalier Universitaire de Québec - Université Laval Research Center, Québec City, QC, Canada. 157Department of Clinical Pathology The University of Melbourne, Melbourne, VIC, Australia. 158Population Oncology BC Cancer, Vancouver, BC, Canada. 159Department of Population and Public Health University of British Columbia, Vancouver, BC, Canada. 160Clinical Genetics Research Lab, Department of Cancer Biology and Genetics Memorial Sloan-Kettering Cancer Center, New York, NY, USA. 161The Curtin UWA Centre for Genetic Origins of Health and Disease Curtin University and University of Western Australia, Perth, Western Australia, Australia. 162Service de Génétique Institut Curie, Paris, France. 163Department of Tumour Biology INSERM U830, Paris, France. 164Université Paris Descartes, Paris, France. 165Division of Breast Cancer Research Institute of Cancer Research, London, UK. 166Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA. 167Epigenetic and Stem Cell Biology Laboratory National Institute of Environmental Health Sciences, NIH Research Triangle Park, Triangle Park, NC, USA. 168Hereditary Cancer Program ONCOCBELL-IDIBELL-IDIBII-IGPT, Catalan Institute of Oncology, CIBERONC, Barcelona, Spain. 169Department of Medicine Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. 170Program in Cancer Genetics, Departments of Human Genetics and Oncology McGill University, Montréal, QC, Canada. 171Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Center, University of Cambridge, Cambridge, UK. 172Department of Cancer Biology and Genetics The Ohio State University, Columbus, OH, USA. 173Institute of Human Genetics Pontificia Universidad Javeriana, Bogota, Colombia. 174Department of medicine University Of Melbourne, Melbourne, VIC, Australia. 175Department of Medical Oncology Beth Israel Deaconess Medical Center, Boston, MA, USA. 176Department of Health Genomics, Division of Epidemiology Mayo Clinic, Rochester, MN, USA. 177Fundación Pública Gallega Medicina Xenómica-SERGAS, Instituto de Investigación Sanitaria Santiago de Compostela (IDIS); CIBERER, Santiago de Compostela, Spain. 178Biostatistics and Computational Biology Branch National Institute of Environmental Health Sciences, NIH Research Triangle Park, Triangle Park, NC, USA. 179Clinical Cancer Genomics City of Hope, Duarte, CA, USA. 180Department of Surgical Sciences Uppsala University, Uppsala, Sweden. 181Department of Oncology Mayo Clinic, Rochester, MN, USA. 182Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center Vanderbilt University School of Medicine, Nashville, TN, USA. 183Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. 184Department of Preventive Medicine Seoul National University College of Medicine, Seoul, Korea. 185Department of Biomedical Sciences Seoul National University Graduate School, Seoul, Korea. 186Cancer Research Institute Seoul National University, Seoul, Korea. 187Department of Clinical Genetics Odense University Hospital, Odense, Denmark. 188Department of Laboratory Medicine and Pathology Mayo Clinic, Rochester, MN, USA. 277These authors jointly supervised this work: Douglas F. Easton, Nadine Andrieu, Antonis C. Antoniou. *Lists of authors and their affiliations appear at the end of the paper. Email: nadine.andrieu@cune.fr; aca20@medschl.cam.ac.uk

**GEMO Study Collaborators**

Ophélie Bertrand162, Sandrine Caputo162, Anaïs Dupré162, Marine Le Montec162, Dominique Stoppa-Lyonnet162,163,164, Muriel Belott163, Anne-Marie Birot163, Bruno Bucher163, Emmanuelle Fournier163, Marion Gauthier-Villars163, Lisa Golmard163, Claude Houdayer163, Virginie Moncoutier163, Antoine de Pauw163, Claire Saule163, Fabienne Lesueur163,12,3,5, Noura Melbrouk12,3,5, Olga Sinilnikova189, Sylvie Mazoyer189, Francesca Damiola189, Laure Barjhouy189, Carole Verny-Pierre189, Mélanie Léone189, Nadia Bouthry-Kryza189, Alain Calender189, Sophie Giraud189, Olivier Caron190, Marine Guillaud-Bataille190, Brigitte Bressac-de-Paillerets190, Yves- Jean Bignon191, Nancy Uhrhammer191, Christine Lasset192, Valérie Bonadona192, Pascaline Berthet193, Dominique Vau192,193, Laurent Castera193, Tetsuro Noguchi194, Cornel Popovici194, Hagay Sobo194, Viola Bourdon194, Tetsuro Noguchi194, Audrey Remenieras194, Catherine Noguès194, Isabelle Coupiet195, Pascal Pujol195, Aurélie Dumont196,
François Révillion196, Claude Adenis196, Danièle Muller197, Emmanuelle Barouk-Simonet198, Françoise Bonnet199, Virginie Bubien198, Nicolas Sevenet198, Michel Longy199, Christine Toulas199, Rosine Guimbaud199, Laurence Gladieff199, Viviane Feille199, Dominique Leroux200, Hélène Dreyfus200, Christine Rebischung200, Magalie Peysselon200, Fanny Coron201, Laurence Faivre201, Amandine Baurand201, Caroline Jacquot201, Geoffrey Bertolone201, Sarab Lizard201, Fabienne Prieur202, Marine Lebrun202, Caroline Kientz202, Sandra Fert Ferrer203, Véronique Mar204, Laurence Vénat-Bouvet205, Capucine Delnâte206, Stéphane Bézieur206, Isabelle Mortemousse207,208, Florence Coulet209, Chrystelle Colas209, Florent Soubrier209, Mathilde Warcoing209, Johanna Sokolowska210, Myriam Bronne210, Marie-Agnès Collonge-Rame211, Alexandre Damette211, Paul Gesta212,213, Hakima Lalloua214, Jean Chiesa215, Denise Molina-Gomes216 & Olivier Ingster217

189Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon - Centre Léon Bérard, Lyon, France. 190Institut Gustave Roussy, Villejuif, France. 191Centre Jean Perrin, Clermont-Ferrand, France. 192Centre Léon Bérard, Lyon, France. 193Centre François Baclesse, Caen, France. 194Institut Paoli-Calmettes, Marseille, France. 195CHU Arnaud-de-Villeneuve, Montpellier, France. 196Centre Oscar Lambret, Lille, France. 197Centre Paul Strauss, Strasbourg, France. 198Institut Bergonié, Bordeaux, France. 199Institut Claudius Regaud, Toulouse, France. 200CHU, Grenoble, France. 201CHU, Dijon, France. 202CHU, St-Etienne, France. 203Hôpital Dieu Centre Hospitalier, Chambéry, France. 204Centre Antoine Lacassagne, Nice, France. 205CHU, Limoges, France. 206CHU, Nantes, France. 207CHU Bretonneau, Tours, France. 208Centre Hospitalier de, Bourges, France. 209Groupe Hospitalier Pitié-Salpétrière, Paris, France. 210CHU Vandoeuvre-les-Nancy, France. 211CHU, Besançon, France. 212CHU Poitiers, Centre Hospitalier d’Angoulême, Poitiers, France. 213Centre Hospitalier de Niort, Niort, France. 214Centre Hospitalier de La Rochelle, La Rochelle, France. 215CHU Nîmes Carémeau, Nîmes, France. 216CHU, Poissy, France. 217CHU, Angers, France.

EMBRACE Collaborators

Helen Gregory218, Zosia Miedzybrodzka218, Patrick J. Morrison219, Kai-ren Ong220, Alan Donaldson221, Marc Tischkowitz217,219, Mark T. Rogers222, M. John Kennedy223, Mary E. Porteous224, Carole Brewer225, Rosemarie Davidson226, Louise Izatt227, Angela Brady228, Julian Barwell229, Julian Adlard230, Claire Foo231, D. Gareth Evans67,68, Fiona Lalloo232, Lucy E. Side233, Jacqueline Eason234, Alex Henderson235, Lisa Walker236, Rosalind A. Eeles237, Jackie Cook238, Katie Snape239, Diana Eccles63, Alex Murray240 & Emma McCann241

218North of Scotland Regional Genetics Service, NHS Grampian & University of Aberdeen, Foresterhill, Aberdeen, UK. 219Northern Ireland Regional Genetics Centre, Belfast Health and Social Care Trust, and Department of Medical Genetics, Queens University Belfast, Belfast, UK. 220West Midlands Regional Genetics Service, Birmingham Women’s Hospital Healthcare NHS Trust, Edgbaston, Birmingham, UK. 221Clinical Genetics Department, St Michael’s Hospital, Bristol, UK. 222All Wales Medical Genetics Services, University Hospital of Wales, Cardiff, UK. 223Academic Unit of Clinical and Molecular Oncology, Trinity College Dublin and St James’s Hospital, Dublin, Eire. 224South East of Scotland Regional Genetics Service, Western General Hospital, Edinburgh, UK. 225Department of Clinical Genetics, Royal Devon & Exeter Hospital, Exeter, UK. 226Clinical Genetics, Southern General Hospital, Glasgow, UK. 227Clinical Genetics, Guy’s and St. Thomas’ NHS Foundation Trust, London, UK. 228North West Thames Regional Genetics Service, Kennedy-Galton Centre, Harrow, UK. 229Leicestershire Clinical Genetics Service, University Hospitals of Leicester NHS Trust, Leicester, UK. 230Yorkshire Regional Genetics Services, Leeds, UK. 231Department of Clinical Genetics, Alder Hey Hospital, Eaton Road, Liverpool, UK. 232Genetic Medicine, Manchester Academic Health Sciences Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK. 233North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Trust, London, UK. 234Nottingham Clinical Genetics Service, Nottingham University Hospitals NHS Trust, Nottingham, UK. 235Institute of Genetic Medicine, Centre for Life, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK. 236Oxford Regional Genetics Service, Churchill Hospital, Oxford, UK. 237Oncogenetics Team, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, UK. 238Sheffield Clinical Genetics Service, Sheffield Children’s Regional Hospital, Sheffield, UK. 239South West Thames Regional Genetics Service, St. Georges Hospital, Cranmer Terrace, Tooting, London, UK. 240All Wales Medical Genetics Services, Singleton Hospital, Swansea, UK. 241All Wales Medical Genetics Service, Glen Clwyd Hospital, Rhyl, UK.

KConFab Investigators

Stephen Fox242, Ian Campbell242, Georgia Chenevix-Trench7, Amanda Spurde7, Penny Webb7, Anna de Fazio243, Margaret Tassell244, Judy Kirk245, Geoff Lindeman246, Melanie Price247, Melissa Southey85,157, Roger Milne83,84,85, Sid Deb248 & David Bowtell249

242Peter MacCallum Cancer Centre, Melbourne, Australia. 243Westmead Millennium Institute, Sydney, Australia. 244BCNA delegate, Community Representative, Melbourne, Australia. 245Westmead Hospital, Sydney, Australia. 246Walter and Eliza Hall Institute, Melbourne, Australia. 247University of Sydney, Sydney, Australia. 248Melbourne Health, Melbourne, Australia. 249Garvan Institute of Medical Research, Sydney, Australia.
HEBON Investigators

Annemieke H. van der Hout\textsuperscript{250}, Ans M. W. van den Ouweland\textsuperscript{251}, Arjen R. Mensenkamp\textsuperscript{252}, Carolien H. M. van Deurzen\textsuperscript{253}, Carolien M. Kets\textsuperscript{252}, Caroline Seynaeve\textsuperscript{101}, Christi J. van Asperen\textsuperscript{254}, Cora M. Aalfs\textsuperscript{255}, Encarna B. Gómez Garcia\textsuperscript{256}, Flora E. van Leeuwen\textsuperscript{257}, Frans B. L. Hogervorst\textsuperscript{9}, G. H. de Bock\textsuperscript{258}, Hanne E. J. Meijers-Heijboer\textsuperscript{259}, Inge M. Obdeijn\textsuperscript{260}, J. Margriet Collée\textsuperscript{251}, J. J. P. Gille\textsuperscript{259}, Jan C. Oosterwijk\textsuperscript{250}, Juul T. Wijnen\textsuperscript{56,254}, Lizet E. van der Kolk\textsuperscript{9}, Maartje J. Hooning\textsuperscript{101}, Margreet G. E. M. Ausems\textsuperscript{261}, Marian J. E. Mourits\textsuperscript{262}, Marinus J. Blok\textsuperscript{263}, Marjanka K. Schmidt\textsuperscript{118,153}, Matti A. Rookus\textsuperscript{257}, Muriel A. Adank\textsuperscript{259}, Peter Devilee\textsuperscript{55,56}, Rob B. van der Luijt\textsuperscript{261}, T. C. T. E. van Cronenburg\textsuperscript{254}, Carmen C. van der Pol\textsuperscript{264}, Nicola S. Russell\textsuperscript{265}, Sabine Siesling\textsuperscript{266}, Lucy Overbeek\textsuperscript{267}, R. Wijnands\textsuperscript{257} & Judith L. de Lange\textsuperscript{257}

\textsuperscript{250}Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands. \textsuperscript{251}Department of Clinical Genetics, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands. \textsuperscript{252}Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. \textsuperscript{253}Department of Pathology, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands. \textsuperscript{254}Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. \textsuperscript{255}Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands. \textsuperscript{256}Department of Clinical Genetics, MUMC, Maastricht, The Netherlands. \textsuperscript{257}Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands. \textsuperscript{258}Department of Oncological Epidemiology, University Medical Center, Groningen University, Groningen, The Netherlands. \textsuperscript{259}Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands. \textsuperscript{260}Department of Radiology, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands. \textsuperscript{261}Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. \textsuperscript{262}Department of Gynaecological Oncology, University Medical Center, Groningen University, Groningen, The Netherlands. \textsuperscript{263}Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands. \textsuperscript{264}Department of Oncological and Endocrine Surgery, University Medical Center Utrecht, Utrecht, The Netherlands. \textsuperscript{265}Department of Radiotherapy, Netherlands Cancer Institute, Amsterdam, The Netherlands. \textsuperscript{266}The Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands. \textsuperscript{267}Foundation PALGA (The Nationwide Network and Registry of Histo- and Cytopathology in the Netherlands), Houten, The Netherlands.

ABCTB Investigators

Christine Clarke\textsuperscript{51}, Dinny Graham\textsuperscript{51}, Mythily Sachchithananthan\textsuperscript{51}, Deborah Marsh\textsuperscript{268}, Rodney Scott\textsuperscript{269}, Robert Baxter\textsuperscript{270}, Desmond Yip\textsuperscript{271}, Jane Carpenter\textsuperscript{272}, Alison Davis\textsuperscript{273}, Nirmala Pathmanathan\textsuperscript{274,275} & Peter Simpson\textsuperscript{276}

\textsuperscript{268}University of Technology Sydney, Translational Oncology Group, School of Life Sciences, Faculty of Science, Ultimo, NSW, Australia. \textsuperscript{269}School of Biomedical Sciences, University of Newcastle, Newcastle; Hunter Medical Research Institute and NSW Health Pathology North, Newcastle, Australia. \textsuperscript{270}Kolling Institute of Medical Research, University of Sydney, St Leonards, NSW, Australia. \textsuperscript{271}Department of Medical Oncology, The Canberra Hospital, Canberra, ACT, Australia. \textsuperscript{272}Scientific Platforms, The Westmead Institute for Medical Research, The University of Sydney, Sydney, NSW, Australia. \textsuperscript{273}The Canberra Hospital, Garran, ACT; The Australian National University, Canberra, ACT, Australia. \textsuperscript{274}Westmead Breast Cancer Institute, Western Sydney Local Health District, Westmead, New South Wales, Australia. \textsuperscript{275}University of Sydney, Western Clinical School, Westmead, New South Wales, Australia. \textsuperscript{276}UQ Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Herston, QLD, Australia.