

Copy number variants as modifiers of breast cancer risk for *BRCA1/BRCA2* pathogenic variant carriers

The contribution of germline copy number variants (CNVs) to risk of developing cancer in individuals with pathogenic *BRCA1* or *BRCA2* variants remains relatively unknown. We conducted the largest genome-wide analysis of CNVs in 15,342 *BRCA1* and 10,740 *BRCA2* pathogenic variant carriers. We used these results to prioritise a candidate breast cancer risk-modifier gene for laboratory analysis and biological validation. Notably, the HR for deletions in *BRCA1* suggested an elevated breast cancer risk estimate (hazard ratio (HR) = 1.21), 95% confidence interval (95% CI = 1.09–1.35) compared with non-CNV pathogenic variants. In contrast, deletions overlapping *SULT1A1* suggested a decreased breast cancer risk (HR = 0.73, 95% CI 0.59–0.91) in *BRCA1* pathogenic variant carriers. Functional analyses of *SULT1A1* showed that reduced mRNA expression in pathogenic *BRCA1* variant cells was associated with reduced cellular proliferation and reduced DNA damage after treatment with DNA damaging agents. These data provide evidence that deleterious variants in *BRCA1* plus *SULT1A1* deletions contribute to variable breast cancer risk in *BRCA1* carriers.

Women who carry pathogenic variants in *BRCA1* (OMIM 113705) and *BRCA2* (OMIM 600185) have greatly increased risk of developing breast cancer. Recent cumulative risk estimates in high-risk families for developing breast cancer by age 80 years were 72% (95% confidence interval (95% CI, 65–79%)) for *BRCA1* and 69% (95% CI, 61–77%) for *BRCA2* pathogenic variant carriers¹. The significant variation of age at diagnosis of breast cancer between pathogenic variant carriers suggests additional factors, such as common inherited genetic variants, influence disease penetrance^{2,3}. Large genome-wide association studies, facilitated by the Consortium of Investigators of Modifiers of *BRCA1/BRCA2* (CIMBA^{2,3}), have demonstrated that >60 single nucleotide polymorphisms (SNPs) or small insertions or deletions (Indels) associated with cancer risk in the general population also are associated with breast cancer risk for *BRCA1/2* pathogenic variant carriers^{4–6}. Moreover, population-based breast cancer polygenic risk scores are associated with modified breast cancer risk for carriers^{7,8}. However, the identified single nucleotide variant modifiers account for <10% of heritable variation in risk in *BRCA1/2* pathogenic variant carriers⁵.

Copy number variants (CNVs) cover between 5–10% of the human genome and, based on nucleotide coverage, are responsible for the majority of variation in the human genome^{9,10}. CNVs exhibit substantial variability in both size and frequency and can disrupt gene function significantly by altering gene dosage, coding sequences, and gene regulation¹¹. Germline CNVs overlapping the *BRCA1* and *BRCA2* gene loci are associated with the pathogenesis of breast cancer, accounting for <5% of known pathogenic variants in these genes (<https://www.ncbi.nlm.nih.gov/clinvar>). Multiple studies have utilised a genome-wide approach to identify associations between rare and common CNVs and the risk of developing breast cancer^{12–16}.

CNVs previously have been shown to be modifiers of hereditary breast cancer risk. In a genome-wide association study (GWAS) of CNVs in 2500 *BRCA1* pathogenic variant carriers, 52 gene loci were associated (unadjusted $p < 0.05$) with breast cancer risk¹⁴. Although no variant reached the widely-adopted genome-wide statistical significance threshold applied for SNP-centric GWAS ($p < 5 \times 10^{-8}$) and the study sample size was relatively small, the specific genes disrupted by CNVs had plausible biological consequences regarding cancer development. These data suggested that CNVs are an important modifier of hereditary breast cancer risk and highlighted the need for larger and more comprehensive CNV studies.

In this study, we conducted genome-wide CNV analyses of 15,342 *BRCA1* and 10,740 *BRCA2* pathogenic variant carriers using genotype data generated by the OncoArray Network¹⁷. We applied in silico and in vitro analyses to characterise a novel risk association between deletions overlapping *SULT1A1* (OMIM 171150) and decreased breast cancer risk for *BRCA1* pathogenic variant carriers.

Results

Copy number variants. 857,647 CNVs (327,530 deletions and 530,117 duplications) were called in study participants, of which 374,210 CNVs (43.6%; 136,534 deletions and 237,676 duplications) overlapped at least one of 16,395 different gene regions. On average, each genome carried 14.4 CNVs (range = 0–63) that overlapped an average of 21.7 genes (range = 0–236). On average, duplications were detected nearly twice as often as deletions (9.1 (range = 0–55) versus 5.2 (range = 0–58), respectively) and affected twice as many gene regions (14.6 (range = 0–220) versus 7.0 (range = 0–207), respectively).

Evaluation of CNV calling. The sensitivity and specificity of our CNV calling was assessed by comparing diagnostically identified

BRCA1 and *BRCA2* CNVs to the CNVs called by our analysis. In our cohort, 1,138 *BRCA1* and 166 *BRCA2* diagnostically identified CNVs overlapped five or more probes and passed our variant filtering; we detected 678 and 155 for *BRCA1* and *BRCA2*, respectively. Furthermore, our genome-wide analysis called 851 *BRCA1* and 183 *BRCA2* CNVs, of which 151 and 28 CNVs were not supported by diagnostic testing of *BRCA1* and *BRCA2*, respectively. Together, our CNV calling achieved an 82.2% and 84.7% detection specificity and 59.6% and 92.8% detection sensitivity for CNVs in *BRCA1* and *BRCA2*, respectively.

Separate analysis of deletions and duplications found that PennCNV performed better with deletion calling versus duplication calling (Supplementary Data 1). The sensitivity of calling *BRCA1* and *BRCA2* deletions were 84.1% and 91.3%, respectively, while the sensitivity of calling *BRCA1* and *BRCA2* duplications were 68.6% and 39.1%, respectively. Similarly, the specificity of calling *BRCA1* and *BRCA2* deletions was 70.4% and 94.8% and duplications was 25.0% and 60.0%, respectively. A review of the diagnostic CNVs not called by PennCNV found that the majority of uncalled CNVs overlapped were the same variants. For example, 54.1% of *BRCA1* deletions not detected by PennCNV were the same variant (c.5333-36_5406 + 400del510), which was only successfully called by PennCNV for 1.4% of the variant carriers.

Genome-wide CNV calling was assessed further for three cases using whole-genome sequencing (WGS). In the three whole-genome sequenced cases, 4444 (Case 1), 4540 (Case 2), and 4545 (Case 3) CNV calls passed confidence filtering, of which 1884 (Case 1), 1981 (Case 2), and 1940 (Case 3) overlapped gene regions, respectively (Supplementary Data 2). A total of 10 of 14 (71.4%—Case 1), 23 of 42 (54.8%—Case 2), and 13 of 19 (68.4%—Case 3) of PennCNV calls were supported by a CNV called in the WGS data (Supplementary Data 3). Of the CNVs not supported by WGS, 4 of 4 (100%—Case 1), 12 of 19 (63.2%—Case 2), and 6 of 6 (100%—Case 3), CNVs were supported by a previously-published CNV map¹⁰. All duplications that were not supported by WGS were supported by the CNV map while approximately half the deletions were not (Supplementary Data 4). Each of the three cases also carried a diagnostically identified pathogenic *BRCA1* deletion that was called by WGS CNV calling. Together, these data provide confidence that >80% of deletions and duplications called by PennCNV in this study cohort appear to be true calls.

Prioritization of candidate breast cancer CNV risk loci. To prioritise genes for in silico and functional analyses, we selected candidate gene loci with $p < 0.01$ from retrospective likelihood analysis, effectively restricting hazard ratios to >1.25 and <0.75 (Supplementary Data 5–8). Putative CNVs at 31 gene regions passed this threshold. For 16 of these 31 regions, the proportion of unique CNVs represented in a published human CNV map¹⁰ was <95%. Although none of the CNV regions passed significance thresholds when adjusted for multiple hypothesis testing (See Methods; deletions in *BRCA1* carriers— $p \leq 8 \times 10^{-6}$; duplications in *BRCA1* carriers— $p \leq 5 \times 10^{-6}$; deletions in *BRCA2* carriers— $p \leq 1 \times 10^{-5}$; and duplications in *BRCA2* carriers— $p \leq 6 \times 10^{-6}$), we used these results to prioritise a candidate risk-modifier gene for laboratory analysis and biological validation.

Deletions overlapping *BRCA1* increased breast cancer risk (hazard ratio (HR) = 1.29, 95%CI = 1.13–1.49, $p = 1.98 \times 10^{-4}$) (Supplementary Data 5) for *BRCA1* pathogenic variant carriers. This result was explored further as the analysis did not directly compare the effect of *BRCA1* deletions and *BRCA1* non-deletion pathogenic variants. Clinically diagnosed variants for *BRCA1* and *BRCA2* carriers were categorised by type (deletions, duplications, and small

Table 1 Breast cancer hazard ratio estimates using a single model comparing other *BRCA1/2* pathogenic variants versus (1) deletions, and (2) duplications^a.

	Mutation type	Unaffected	Affected	HR	95%CI		p
<i>BRCA1</i> carriers	Other	8669	8483	1.00	[reference]		
	Deletions	636	789	1.21	1.09	1.35	4.35E-04
	Duplications	153	192	1.21	0.99	1.48	6.60E-02
<i>BRCA2</i> carriers	Other	5913	6204	1.00	[reference]		
	Deletions	85	111	1.11	0.79	1.54	5.54E-01
	Duplications	6	16	1.52	0.61	3.77	3.68E-01

^a Weighted cohort models fitted separately for *BRCA1* and *BRCA2* carriers. Weights were calculated assuming *BRCA1/2* carrier breast cancer incidences from most recent age cohort in Antoniou et al., (2008). Models were stratified by country and Ashkenazi Jewish ancestry, and adjusted for birth cohort and genotyping array. Cluster robust variances were estimated using families as clusters. HR hazard ratio, 95%CI 95% confidence interval; p-value.

Table 2 Putative copy number variants assessed by TaqMan assays.

Gene symbol	CNV type	Variant carrier	Assay ID	Overlap CNV map ¹	Proportion validated
<i>SULT1A1</i>	Deletion	<i>BRCA1</i>	1	Yes	100% (8/8)
<i>TERT</i>	Deletion	<i>BRCA2</i>	1	No	0% (0/1)
			2	No	0% (0/1)
			3	No	33% (1/3)
<i>LSP1</i>	Duplication	<i>BRCA2</i>	1	No	50% (1/2)

CNV copy number variant. ¹Zarrei et al (2015) stringent CNV map.

variants [i.e. nonsense, missense, frame shift, Indel, and splice site]). Assessing the HRs for CNV versus non-CNV pathogenic variants, separately for *BRCA1* and *BRCA2* suggested elevated breast cancer risk for *BRCA1* deletions (HR = 1.21, 95%CI = 1.09–1.35) but not *BRCA2* deletions (Table 1, Supplementary Data 9). These results remained similar after excluding missense variant carriers from the analysis (Supplementary Data 9). Estimated HRs were elevated for duplications versus non-duplication pathogenic variants (deletions were excluded) for *BRCA1* duplication carriers (HR = 1.21, 95% CI = 0.99–1.48; $p = 0.066$), and *BRCA2* duplication carriers (HR = 1.52, 95%CI = 0.61–3.77, $p = 0.39$); however, results for *BRCA2* were less definitive given the smaller sample size and wide confidence intervals.

Putative duplications overlapping the breast cancer tumour suppressor gene *STK11* suggested decreased risk of breast cancer in our study for both *BRCA1* carriers (HR = 0.49, 95%CI 0.29–0.81, $p = 5.4 \times 10^{-3}$) and *BRCA2* carriers (HR = 0.44, 95% CI 0.22–0.88, $p = 9.2 \times 10^{-3}$). Putative deletions overlapping *TERT* and duplications overlapping *LSP1*, two loci previously shown to be associated with breast cancer risk for *BRCA1* (*TERT* locus) and *BRCA2* (*TERT* and *LSP1* loci) pathogenic variant carriers from SNP-based studies^{18,19}, suggested increased risk (HR = 1.92, 95%CI = 1.06–3.46, $p = 6.0 \times 10^{-3}$) and decreased risk (HR = 0.13, 95%CI = 0.04–0.45, $p = 3.3 \times 10^{-3}$) breast cancer risk for *BRCA2* pathogenic variant carriers in this study, respectively. However, analysis of *TERT* and *LSP1* CNVs using TaqMan assays with available DNA showed only 25% (1/4) of predicted deletions overlapping *TERT* and 50% (1/2) predicted duplications overlapping *LSP1* were successfully validated (Table 2, Supplementary Fig. 1). These results are consistent with the observation that predicted CNVs overlapping *LSP1* and *TERT* were not found in a published genomic map of human CNVs¹⁰.

Identification of *SULT1A1* as a candidate modifier gene. CNV loci suggested to modify breast cancer risk estimates in *BRCA1/2* pathogenic variant carriers, were examined to identify a candidate gene for functional characterisation using in silico and in vitro

assays. *SULT1A1* (sulfotransferase 1A1) was selected as a novel candidate modifier based on potential biological mechanisms of action and because overlapping CNVs had a population frequency above 1%.

In our study, CNV deletions overlapping *SULT1A1* were identified in 1.7% of *BRCA1* pathogenic variant carriers and they suggested a decreased breast cancer HR (HR = 0.73, 95% CI = 0.59–0.91, $p = 9.1 \times 10^{-3}$). Deletions overlapping all eight *SULT1A1* exons of the reference transcript were validated in all eight available DNA samples using both TaqMan assays and multiplex ligation-dependent probe amplification (MLPA; Supplementary Fig. 2), and were identified in the CNV map published by Zarrei et al (Supplementary Fig. 1c). Furthermore, CNVs involving the *SULT1A1* gene locus had an expression dosage effect in breast tumours (Fig. 1).

Characterisation of *SULT1A1* knockdown in *BRCA1*^{+/-} cells.

To model the effect of *SULT1A1* deletions in *BRCA1* carriers, a pair of isogenic MCF7 breast cell lines with and without a pathogenic variant (*BRCA1* c.2432_2433del) were created using CRISPR-Cas9 (Fig. 2a). A comparison of the isogenic control (MCF7-C1) and pathogenic variant carrying cells (MCF7-*BRCA1*^{+/-}) with the parent MCF7 cells (MCF7-WT) showed no significant difference in cell proliferation (Fig. 2b, Supplementary Data 10). There were also no differences in the relative expression of *BRCA1* mRNA between the isogenic MCF7-C1 and MCF7-*BRCA1*^{+/-} lines and the parent MCF7-WT line (Fig. 2c, Supplementary Data 10). However, the MCF7-*BRCA1*^{+/-} cells showed a significant 25% ($p = 4.44 \times 10^{-4}$) reduction in *ESR1* expression (Fig. 2d, Supplementary Data 10) and a significant 78% ($p = 2.03 \times 10^{-3}$) increase in *CYP1A1* expression (Fig. 2e, Supplementary Data 10), consistent with breast cells with a pathogenic *BRCA1* variant²⁰. There was no significant difference in *SULT1A1* mRNA expression between the MCF7-WT and the isogenic MCF7-C1 and MCF7-*BRCA1*^{+/-} cells (Fig. 2f, Supplementary Data 10).

siRNA was used to transiently reduce the relative expression of *SULT1A1* mRNA in the isogenic MCF7-C1 and MCF7-*BRCA1*^{+/-}

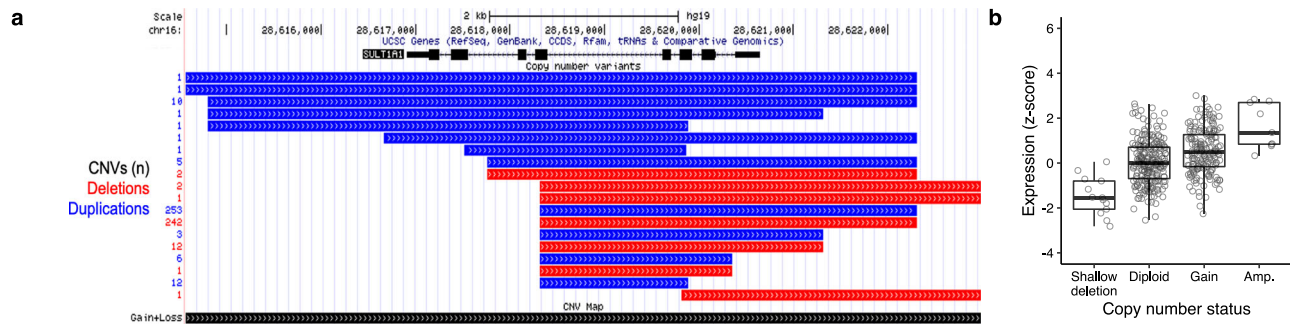


Fig. 1 Characterisation of SULT1A1. **a** Genomic viewer (UCSC) of the SULT1A1 gene locus with copy number variants; deletion (red) and duplication (blue). **b** Dosage effect in breast tumours with SULT1A1 copy number variants. CNVs copy number variants.

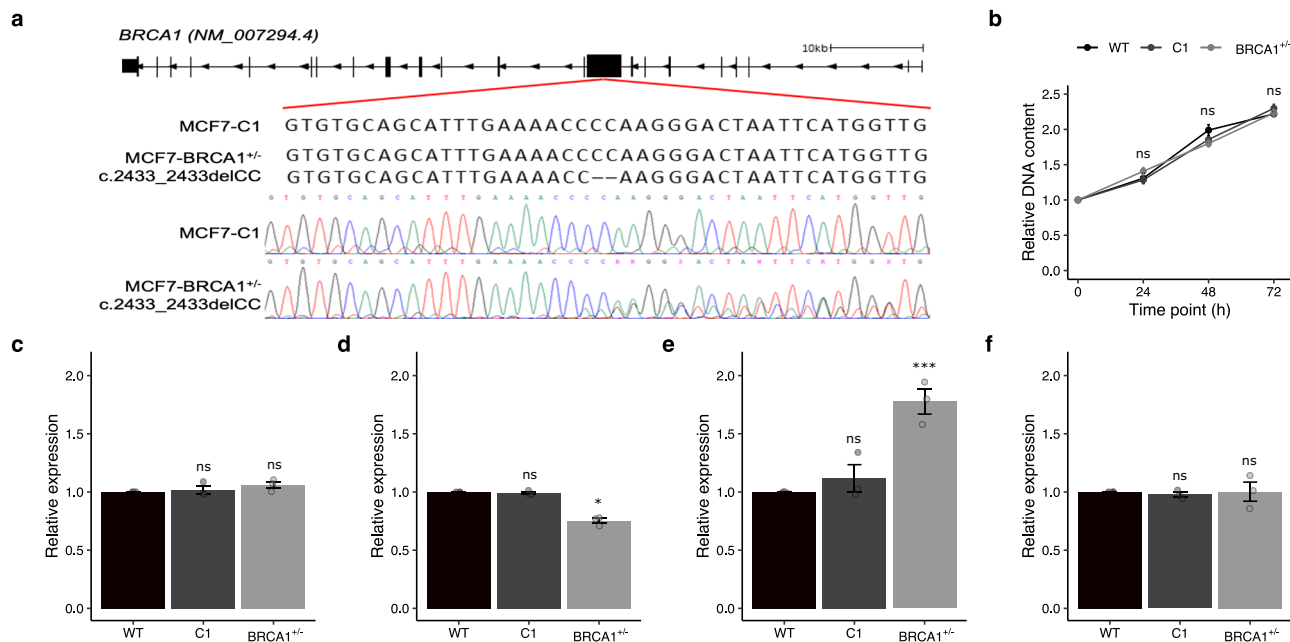


Fig. 2 Characterisation of MCF7-WT, and isogenic MCF7-C1, and MCF7-BRCA1^{+/-} cell lines. **a** Sequence of heterozygous pathogenic BRCA1 c.2432_2433del variant introduced by CRISPR-Cas9. **b** Relative proliferation of MCF7-WT and clonally expanded CRISPR-Cas9 MCF7-C1 and MCF7-BRCA1^{+/-} cells for 72 h post seeding. Relative expression of **c** BRCA1, **d** ESR1, **e** CYP1A1, and **f** SULT1A1 for MCF7-WT, MCF7-C1 and MCF7-BRCA1^{+/-} cells. 4-OHE2 4-hydroxyestradiol, MMC Mitomycin C, Error bars = standard error of the mean; ns = $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; $n = 3$ independent biological replicates; unpaired two-sided t -test.

lines. Compared with the non-targeting siRNA control, the relative expression of *SULT1A1* was approximately half in both isogenic lines 72 h after transfection targeting *SULT1A1* (Fig. 3a, Supplementary Data 11). As deletions overlapping *SULT1A1* were suggested to decrease breast cancer risk in *BRCA1* pathogenic variant carriers, the relative expression of *BRCA1* was quantified to assess if the *SULT1A1* knockdown affected its regulation (Fig. 3b, Supplementary Data 11). However, there was no significant change in the *BRCA1* expression for the *SULT1A1* knockdown cells compared with the transfection control for either of the MCF7-C1 or the MCF7-BRCA1^{+/-} lines.

The proliferation of the transfected MCF7-C1 and MCF7-BRCA1^{+/-} cells was assessed by measuring relative DNA content. Knockdown of *SULT1A1* expression did not alter the proliferation of the MCF7-C1 compared with the transfection control (Fig. 3c, Supplementary Data 11). However, when *SULT1A1* expression was reduced in MCF7-BRCA1^{+/-} cells, there was a 14% ($p = 1.17 \times 10^{-2}$) decrease in proliferation compared with the transfection control after 72 h of growth (Fig. 3d, Supplementary Data 11).

Because *BRCA1*-deficient cells are hypersensitive to DNA damaging agents and have an impaired DNA damage repair

response^{20,21}, we investigated whether knockdown of *SULT1A1* expression altered the amount of damage caused by the DNA damaging agents 4-hydroxyestradiol (4-OHE2) and Mitomycin C (MMC) using the well-established comet and γ H2AX and 53BP1 immunostaining assays. Analysis using comet assays, showed that treatment with 4-OHE2 ($F(1, 8) = 5.79$, $p = 4.28 \times 10^{-2}$; Fig. 3e, Supplementary Data 11) and MMC ($F(1, 8) = 5.73$, $p = 4.36 \times 10^{-2}$; Fig. 3f, Supplementary Data 11) increased comet tail moment length in MCF7-BRCA1^{+/-} cells. There also was evidence that transfection with siSULT1A1 reduced the average comet tail moment length for both 4-OHE2 ($F(1, 8) = 9.76$, $p = 1.41 \times 10^{-2}$; Fig. 3e, Supplementary Data 11) and MMC ($F(1, 8) = 5.69$, $p = 4.43 \times 10^{-2}$; Fig. 3f, Supplementary Data 11) treated MCF7-BRCA1^{+/-} cells. Assessing the effect of *SULT1A1* knockdown in MCF7-BRCA1^{+/-} and MCF7-C1 cells using the γ -H2AX and 53BP1 immunostaining assay gave rise to analogous results (Fig. 3e-h). Although there was evidence that treatment with 4-OHE2 increased the average number of co-localised DNA damage foci ($F(1, 8) = 24.36$, $p = 1.14 \times 10^{-3}$; Fig. 3g, Supplementary Data 11) in MCF7-BRCA1^{+/-} cells, the reduction in the number of foci caused by the siSULT1A1 transfection did not

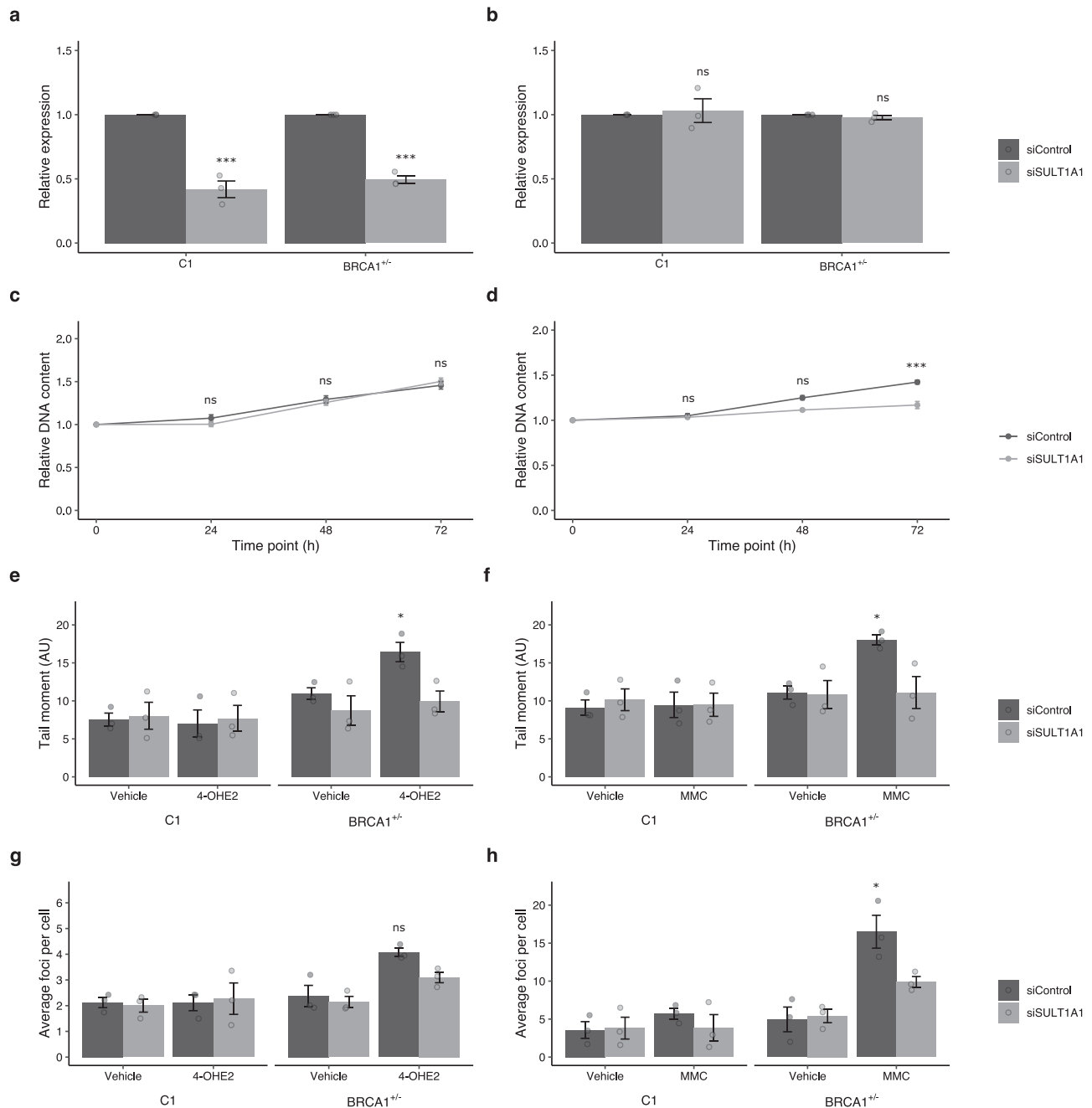


Fig. 3 siRNA knockdown of SULT1A1 reduces proliferation and DNA damage of MCF7-BRCA1^{+/-} cells but not MCF7-C1 cells. Expression of (a) SULT1A1 and (b) BRCA1 72 h post transfection. Relative DNA content of transfected (c) MCF7-C1 and (d) MCF7-BRCA1^{+/-} cells 72 h post transfection. Quantification of transfected DNA damage using the comet assay (e, f) and γ -H2AX/53BP1 foci quantification (g, h) for MCF7-C1 and MCF7-BRCA1^{+/-} cells, with and without siSULT1A1 transfection, 21 h post treatment with 1 μ m 4-OHE2 (e and g) or 10 μ m MMC (f and h). Significance of differences in relative expression and DNA content was determined by unpaired two-sided *t*-test. Gene expression and DNA content of siSULT1A1 transfected cells normalised to siControl transfected cells. Differences in DNA damage were determined by two-way analysis of variance. Example images of comet and γ -H2AX/53BP1 foci shown in Supplementary Figs. 3–6. 4-OHE2 = 4-hydroxyestradiol; MMC = Mitomycin C; Error bars = standard error of the mean; ns = $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; $n = 3$ independent biological replicates.

reach statistical significance ($F(1, 8) = 5.02$, $p = 5.5 \times 10^{-2}$). Additionally, for the MMC treated MCF7-BRCA1^{+/-} cells there was a significant interaction between the siSULT1A1 transfection and MMC treatment ($F(1, 8) = 5.80$, $p = 4.3 \times 10^{-2}$; Fig. 3h, Supplementary Data 11). This effect was significant for MMC treated MCF7-BRCA1^{+/-} cells ($F(1, 4) = 8.44$, $p = 4.4 \times 10^{-2}$) but not the vehicle control ($F(1, 4) = 0.06$, $p = 8.19 \times 10^{-1}$). There was no evidence that siRNA transfection or drug

treatments affected the average comet tail moment or the number of γ -H2AX and 53BP1 foci for the MCF7-C1 cells. Example images of comet and γ -H2AX and 53BP1 immunostaining assay are shown in Supplementary Figs. 3–6.

Discussion

Germline CNVs are an important source of genetic variation that have previously been understudied in relation to breast and

ovarian cancer risk. Here, we have conducted the largest and most comprehensive genome-wide association study of CNVs and breast cancer risk for *BRCA1* and *BRCA2* pathogenic variant carriers. We identified putative CNVs in up to 31 putative gene regions that were associated (unadjusted $P < 0.01$) with breast cancer risk for *BRCA1/2* pathogenic variant carriers, with CNVs at 15 of these regions present in a human CNV map¹⁰. Although none of the CNV regions passed significance thresholds when adjusted for multiple hypothesis testing, we used these results to prioritise a candidate risk-modifier gene for laboratory analysis and biological validation. Consistent with observations from the human CNV map, we validated positive CNV calls overlapping the *SULT1A1* gene, and revealed false positive CNV calls at two candidate modifier gene regions (*LSP1* and *TERT*). CNV deletions overlapping the lead candidate modifier *SULT1A1* showed decreased breast cancer risk in *BRCA1* pathogenic variant carriers. In silico analysis of *SULT1A1* suggested that deletions overlapping this gene leads to reduced expression. In vitro analyses showed that reduced *SULT1A1* expression in cells carrying a heterozygous *BRCA1* pathogenic variant led to reduced cellular proliferation and reduced DNA damage after treatment with DNA damaging agents.

Both SNP and CNV variants at the *SULT1A1* locus have previously been shown to be associated with *SULT1A1* enzymatic activity. The common *SULT1A1* p.(Arg213His) (rs9282861) polymorphism leading to the *SULT1A1**2 variant has been examined in a series of functional and association studies. *SULT1A1**2 has a two-fold lower catalytic activity and stability than its high-activity p.Arg213 counterpart (*SULT1A1**1), and has been associated with increased cancer risk in multiple tissue types^{22–25}. Studies examining the association between the rs9282861 polymorphism and breast cancer risk have yielded inconsistent results showing an increase in risk in some studies^{26–29} and no association in others^{30–32}. The OncoArray probe for rs9282861 failed quality control, therefore no genotype data was available. Interestingly, the rs200802208 Indel located near *SULT1A1* was imputed from these arrays and analyses showed that the del allele was associated with decreased risk of breast cancer in *BRCA1* pathogenic variant carriers (HR = 0.48, 95%CI 0.29–0.79, $p = 4.3 \times 10^{-3}$)⁵. rs9282861 is not in 1000 G reference panel, therefore it is unknown whether this SNP exists in linkage disequilibrium with rs200802208. CNVs overlapping *SULT1A1* are strongly associated with *SULT1A1* activity and explain more of the observed in vitro variability in *SULT1A1* activity than SNPs, with activity proportional to *SULT1A1* copy number^{33–36}. Individuals who are homozygous null for *SULT1A1* do not present with any overt phenotypes³⁷. This finding corresponds with results from phenotypic analyses of mouse *SULT1A1* knockouts which are viable, and which also lack any outward phenotype^{38,39}. However, the absence of functional *SULT1A1* enzyme in mouse knockouts has been reported to reduce the number of DNA adducts caused by DNA damaging agents that are converted to mutagenic metabolites by *SULT1A1*³⁹.

The mechanism by which CNV deletions overlapping *SULT1A1* were associated with lower *BRCA1*-associated breast cancer risk may be linked to the production of potentially toxic catechol oestrogens by the Cytochromes P450 (CYP) enzymes. *SULT1A1* is an important SULT isoform that is expressed widely in human tissues and plays an important role in the metabolism, bioactivation, and detoxification of carcinogens, medications, and steroid hormones^{33,34}. *SULT1A1* has established germline common CNVs and SNPs that are known to alter its activity⁴⁰. *SULT1A1* is most abundantly expressed in the liver, but is also expressed in the brain, breast, intestine, and endometrium^{41–44}. *SULT1A1* expression is related to disease state, with plentiful expression in most breast tumours^{42,45}. In normal breast cells,

BRCA1 regulates oestrogen metabolism and metabolite-mediated DNA damage by repressing transcription of the oestrogen-metabolising enzyme CYP1A1^{20,46}. However, levels of CYP1A1 are higher in breast cells lacking *BRCA1* function and promote the formation of the carcinogenic 2-hydroxyestradiol (2-OHE2)²⁰. Further metabolism of 2-OHE2 by catechol O-methyltransferase (COMT) is thought to have a risk reducing effect by catalysing the formation of 2-methoxyestradiol (2-MeOE2), a metabolite which interacts with the tubulin colchicine-binding site during polymerisation and which has anticarcinogenic effects by suppressing cell proliferation. In turn, *SULT1A1* is an efficient catalyst of 2-MeOE2 sulfation producing 2-MeOE2-3S, a sulfate conjugate with diminished activity^{33,47}. It is possible that the decreased risk associated with *SULT1A1* deletions for *BRCA1* pathogenic variant carriers and the decrease in cell proliferation and amount of DNA damage for MCF7-*BRCA1*^{+/-} with a *SULT1A1* knockdown cells may both be linked to 2-MeOE2 abundance. That is, reduced *SULT1A1* activity promotes the accumulation of 2-MeOE2 and slows the proliferation of breast cells with unbalanced E2 metabolism. Indeed, the *SULT1A1* substrate, 2-MeOE2, has previously been proposed as a potential preventative agent for breast cancer⁴⁸. A similar relationship between CYP1A1 and *SULT1A1* activity and reduced breast cancer risk has been demonstrated previously. In a study of pairwise combinations of oestrogen metabolism alleles and breast cancer risk, the *SULT1A1**2 genotype was assessed in combination with a *CYP1A1* missense variant (*CYP1A1**2 C) that has increased inducibility to produce catechol oestrogens⁴⁹. For European-Americans, carrying the *CYP1A1**2 C genotype was associated with increased breast cancer risk (odds ratio (OR) = 1.71, 95%CI = 1.09–2.67). However, carrying the *CYP1A1**2 C allele in combination with a *SULT1A1**2 allele was strongly protective against developing breast cancer (OR = 0.14, 95% CI = 0.04–0.56) compared with women carrying only the *CYP1A1**2 C allele. There was no association between the *CYP1A1**2 C (rs1048943) polymorphism and breast cancer risk in *BRCA1* pathogenic variant carriers⁵. These results further suggest that the balance between the generation of catecholestrogens and catecholesterogen sulfation may be an important mechanism for modulating breast cancer risk and worthy of future investigation.

Our study provides strong evidence that deletions overlapping *BRCA1* are associated with a 1.21-fold higher risk of developing breast cancer. Large deletions in *BRCA1* have previously been shown to be associated with an increased risk of breast cancer risk (OR = 1.42) compared with carriers of *BRCA1* pathogenic single nucleotide variants or Indels⁵⁰. Similarly, a series of studies have reported a higher incidence of CNVs in both *BRCA1* and *BRCA2* when cases have a family history that includes high-risk features, such as early-onset disease^{51,52}. Although a mechanism that explains the higher risk for *BRCA1* deletion carriers is unclear, one possible explanation is that large genomic variants disrupt key *BRCA1* domains or cause nonsense-mediated mRNA decay, whereas some single nucleotide variants in *BRCA1* avoid nonsense-mediated decay and retain partial function. For example, the p.(Arg1699Gln) variant in *BRCA1* produces a protein with ambiguous behaviour in a variety of functional assays and is also associated with an intermediate risk^{53,54}. Furthermore, variants in both *BRCA1* and *BRCA2* are proposed to have variant-specific risks, that coincide with known or hypothesised functional domains and vary by variant type and location^{1,55}. Our results support the hypothesis that breast cancer risk for women carrying large deletions in *BRCA1* is greater than those pathogenic single nucleotide variants or Indels, which may have implications for clinical risk assessment and management of CNV carriers.

Despite this being the largest sample size of *BRCA1* and *BRCA2* pathogenic variant carriers available to date, the low frequency of CNVs results in limited power for detecting significant associations after adjusting for multiple comparisons. As a result, a nominal screening threshold of 0.01 was used which is arbitrary and is therefore a limitation of the study. Nevertheless, this is the largest extant dataset available for examining genetic modifiers of *BRCA1* and *BRCA2* related risk. While larger studies, such as the new Confluence project (<https://dceg.cancer.gov/research/cancer-types/breast-cancer/confluence-project>), may lead to improved statistical power to detect CNV associations, evaluating uncommon genetic variation such as CNVs that overlap *SULT1A1* and other potential modifier genes in *BRCA1/2* pathogenic variant carriers will remain a challenge. Furthermore, CNV calling algorithms have limitations which lead to false CNV calls, thus highlighting the importance of using ancillary data to prioritise regions for downstream analyses. Here we show that functional analysis of a candidate modifier gene using a model cell line is able to provide additional evidence that *SULT1A1* deletions lead to reduced risk of breast cancer in *BRCA1* pathogenic variant carriers. If verified, future therapeutic intervention studies targeting *SULT1A1* in *BRCA1* pathogenic variant carriers may lead to new medical options for reducing breast cancer risk.

In conclusion, our study provides evidence that CNVs contribute to the variability in breast cancer risk among *BRCA1* and *BRCA2* pathogenic variant carriers. Characterising pathogenic variant type in *BRCA1*, and future screening for deletions overlapping *SULT1A1*, may produce variables to be incorporated with other modifying factors to develop a more comprehensive model of breast cancer risk. For example, integrating these genetic data into the CanRisk Web Tool (<https://www.canrisk.org/>)⁵⁶ along with family history, lifestyle/hormonal risk factors, common genetic susceptibility variants, and mammographic density, may further improve breast cancer risk predictions. Such a model may better inform patient decisions regarding breast cancer risk management.

Methods

Study cohort. Female *BRCA1* and *BRCA2* pathogenic variant carriers were from study centres across North America, Europe, and Australia participating in CIMBA (Supplementary Data 12), as reported previously^{4,5}. Eligibility criteria for study participants included: (1) female carriers of *BRCA1* or *BRCA2* pathogenic variants; and (2) minimum 18 years of age at recruitment. A complete list of *BRCA1* and *BRCA2* pathogenic variants are deposited in the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/submitters/505954/>) and a filtered list of those in participants that were analysed in this study (post-quality control) is shown in Supplementary Data 13. There were 7725 (50.4%) *BRCA1* and 5488 (51.1%) *BRCA2* pathogenic variant carriers diagnosed with breast cancer (Supplementary Data 14). All participants were recruited for research studies using ethically approved protocols at host institutions.

CNV detection and quality control. DNA samples were genotyped using the OncoArray-500k BeadChip (Illumina) with 533,631 probes, and standard sample quality control exclusions were performed as previously described for the SNP genotype analysis¹⁷. GenomeStudio (Illumina) was used to export Log R Ratios (LRRs) and B allele frequencies (BAFs) for each sample as previous described¹⁴. A principal components adjustment (PCA) was run on the LRR to remove noise using the bigca package (V1.1)⁵⁷ in the statistical platform R (V3.5.2)⁵⁸. CNV calls were generated using PennCNV⁵⁹. Probes that failed to cluster using Illumina's GenTrain algorithm ($n = 4857$) and probes on the Y chromosome were removed from these results. Neighbouring CNVs with a gap of <20% of the total length of the combined CNVs, were merged using the PennCNV `clean_cnv.pl` script. For the current study we determined genetic ancestry using a principal components approach described elsewhere⁵. A total of 15679 *BRCA1* and 10981 *BRCA2* pathogenic variant carriers of European ancestry were assessed.

The study cohort was filtered to remove samples that failed study requirements or quality controls (Supplementary Fig. 7). Samples were removed if they met criteria listed in Supplementary Fig. 7, or if they met the following criteria: PennCNV measures of LRR standard deviation (s.d.) > 0.28, BAF drift > 0.01, waviness factor deviating from 0 by > 0.05; LRR outliers > 0.1, BAF s.d. \geq 0.2, LRR s.d. \geq 0.4. Additionally, samples with >100 CNVs were excluded. To reduce false

positive calls, only copy number variants called by five or more probes were retained for analysis. A total of 857,647 CNVs carried by 15,342 *BRCA1* and 10,740 *BRCA2* pathogenic variant carriers passed quality control steps and were assessed.

Defining gene-centric CNVs. To identify genomic loci that influence breast cancer risk for *BRCA1* and *BRCA2* pathogenic variant carriers, a non-redundant gene-centric approach was used. Gene regions were derived from the University of California, Santa Cruz (UCSC) GRCh37/Hg19 gene track (updated: 14 June 2013) and were restricted to chromosomes 1–22, and chromosome X. In total, 30,336 gene regions with 27,038 unique gene symbols were derived and used in our analysis. CNVs that overlapped a gene region by one or more base pairs were identified in a genome-wide scan in R (V3.5.3) using the GenomicRanges package (V1.4)⁶⁰. Overall, 374,210 CNVs overlapped one or more of 16,395 unique gene regions and were retained for statistical analysis.

Breast cancer risk association analysis. The association analyses between breast cancer risk and copy number deletions and duplications were conducted separately for *BRCA1* and *BRCA2* pathogenic variant carriers. Study participants were followed from birth until the age at first breast cancer diagnosis, age at ovarian cancer diagnosis or bilateral prophylactic mastectomy (whichever occurred first), or at the age at last observation. Only those diagnosed with breast cancer were considered to be affected. Pathogenic variant carriers with ovarian cancer were considered unaffected, and censored at ovarian cancer diagnosis.

BRCA1 and *BRCA2* pathogenic variant carriers were sampled non-randomly with respect to their disease status. Therefore, to evaluate associations between deletions and duplications and breast cancer risk, we analysed these data using a kinship adjusted score test based on the retrospective likelihood of observing the CNV conditional on the observed phenotype to account for the non-random ascertainment⁶¹. This model is stratified by country and Ashkenazi Jewish ancestry but is unable to adjust for covariates. An approximation method yields HR and 95%CI estimates based on this score test⁶¹. Instances in which a non-overlapping deletion and duplication was called in the same gene region were excluded, however this occurrence was relatively uncommon (<1% of participants were removed after the analysis of 99.3% of gene regions). Retrospective likelihood analysis of variants was performed using R (V3.3.1) and bespoke software (available on request). Conservative significance thresholds were based on the number of effective tests in this gene-centric CNV study. After excluding gene regions with no overlapping CNVs, thresholds were as follows: deletions in *BRCA1* carriers— $p \leq 0.05/6551 = 8 \times 10^{-6}$; duplications in *BRCA1* carriers— $p \leq 0.05/10240 = 5 \times 10^{-6}$; deletions in *BRCA2* carriers— $p \leq 0.05/5094 = 1 \times 10^{-5}$; and duplications in *BRCA2* carriers— $p \leq 0.05/8469 = 6 \times 10^{-6}$. Hypervariable regions of the genome that are prone to CNV calling errors, including the human leucocyte antigens, immunoglobulin superfamily, and olfactory receptor genes were excluded from the final gene lists.

Models to estimate the associations (HRs) of deletions and duplications simultaneously took the form of a weighted cohort analysis^{61,62}. This method assigns different weights to unaffected and affected carriers depending on their age at diagnosis/censoring such that the weighted cohort mimics a true cohort. Weights were calculated using the most recent birth cohort incidence estimates from Antoniou et al.¹⁸. These models were stratified by country and Ashkenazi Jewish ancestry, and further adjusted for genotyping array (iCOGS or OncoArray) birth cohort (<1920, 1920–29, 1930–39, 1940–49, \geq 1950). To account for relatedness, cluster robust variances were estimated using unique family IDs as clusters.

CNV validation. Gene loci found associated with risk were reviewed to identify CNVs for validation using orthogonal technologies. CNVs were prioritised for validation if one or more DNA samples were available, the gene locus was associated with risk (unadjusted p -value < 0.01), and if they overlapped a gene region that had been previously associated with breast cancer risk. Copy number assessment was carried out using TaqMan assays for five different copy number variable regions in three different genes, including one gene (*SULT1A1*) that also was assayed with MLPA. Custom primer and probe sequences and pre-designed assays from Life Technologies used to validate CNVs are listed in Supplementary Data 15. *SULT1A1* MLPA was performed using the *SULT1A1* MLPA kit (C1-0217; MRC-Holland) and analysed using Coffalyser software (v9.4; MRC-Holland), as per the manufacturer's instructions.

Three samples also were evaluated using WGS to assess the genome-wide accuracy of CNV calling. Libraries were prepared for whole-genome sequencing using the KAPA Hyper PCR Free Library preparation kit (V2.1) and 2 \times 150 bp paired end sequenced on the Illumina HiSeqX platform (Kinghorn Centre for Clinical Genomics, Australia). Genomic data were processed using a modified version of the GATK best practise guidelines. Cases were sequenced to an average of 30-fold depth and CNVs were called using Lumpy and CNVnator (Supplementary Data 16). Briefly, FastQ files were generated and adaptors trimmed using Illumina's Bcl2fastq (V2.16). Reads were aligned to the b37d5 (1000 Genomes Project GRCh37 plus decoy) reference genome using BWA-mem (V0.7.10-r789)⁶³, followed by Novosort (V1.03.01) to create coordinated-sorted duplicate marked files. GATK (V3.3) Indel realignment and base quality recalibration were used to create analysis ready reads⁶⁴. Single nucleotide variants

and short insertions and deletions were joint-called using GATK HaplotypeCaller in gVCF mode with variant quality score recalibration. Structural variants, including CNVs, were distinguished from split reads and discordant pairs using lumpy (V0.2.13)⁶⁵ and read depth differences using CNVnator (V0.3.3)⁶⁶.

Dosage effect analysis. Expression and copy number data from the Breast Invasive Carcinoma⁶⁷ datasets were downloaded using cBioPortal⁶⁸. mRNA expression was calculated as a Z-score from all genes and putative copy number alterations were calculated using GISTIC.

Cell culture. The MCF7 breast cancer cell line was purchased from the American Type Culture Collection (ATCC) and maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) foetal bovine serum (FBS). Cells were cultured in a humidified atmosphere of 5% CO₂ at 37 °C, and routinely passaged every 3–4 days. Cells were used up to a maximum of 30 passages.

Development of MCF7-*BRCA1*^{+/-} cell line. MCF7 cells underwent CRISPR-Cas9 editing to create isogenic cells with and without a heterozygous *BRCA1* frameshift variant resulting in premature truncation of the protein (hereafter referred to as a pathogenic variant). The guide RNA was designed to target exon 11 of *BRCA1* and disrupt its function (sequence 5'-GCAGCATTGAAAACCCCAA). MCF7 cells were transfected with plasmid containing gRNA, Cas9 protein, and puromycin resistance (Addgene [ID #62988]—pSpCas9(BB)—2A-Puro (PX459) V2.0). Control cells underwent a parallel transfection protocol with a null guide RNA plasmid. CRISPR-Cas9 treated cells were clonally expanded and the predicted CRISPR cleavage site was amplified by PCR (forward 5'-GAAAGGATCCTGGGTGTTT, reverse 5'-CTTGTTCCCGACTGTGGT) and was Sanger sequenced to identify pathogenic variants. Isogenic lines were cultured in DMEM (1:1; Invitrogen) with 10% (v/v) FBS (Invitrogen) and grown in a humidified atmosphere of 5% CO₂ at 37 °C.

RNA interference. Cells were seeded at 5000 cells per well in 96-well tissue culture plates and allowed to adhere overnight. Cells were transfected with 20 μM of Silencer™ Select siRNA oligonucleotides targeting human *SULT1A1* (s13613, Ambion) or a non-targeting siRNA negative control (Negative Control No. 1 siRNA, Ambion). Cells were transfected using Lipofectamine RNAi max (Invitrogen) according to manufacturer's specifications. After 24 h of transfection media was replaced with normal growth media.

qPCR. Total RNA was extracted 72 h post transfection using the RNAgem-PLUS kit (ZyGem) to assess the level of gene knockdown. cDNA was synthesized using the Superscript III cDNA Synthesis Kit (Invitrogen) and qPCR was performed using Kapa Probe Fast qPCR Master mix (Kapa Biosystems) on the LightCycler 480 (Roche). The 2^{-ΔΔCT} method was used to quantify mRNA expression levels of target genes, where *HPRT1* was used as an internal reference control. Two well-characterised samples from 1000 Genomes Project with known copy number status were used as copy number controls. Gene-specific primers and fluorescent probes are reported in Supplementary Data 17. Statistical significance was assessed by two-tailed Student's *t*-tests between target genes and the siRNA control. Expression differences were considered statistically significant if the *p*-value was <0.05.

Proliferation assay. MCF7 and MCF7-*BRCA1*^{+/-} cells were seeded at 5000 cells per well in 96-well, black walled, clear-bottom tissue culture plates (Greiner). Cells were allowed to adhere overnight before transfection. Media was replaced 24 h post transfection. Forty-eight hours post media change, cell proliferation was assessed using the CyQUANT™ Cell Proliferation Assay Kit (Invitrogen) according to the manufacturer's instructions. Fluorescence was measured on the Varioskan® Flash plate reader (Thermo Fisher Scientific) using a filter combination for excitation at 480 nm and emission at 520 nm.

DNA damage assay. Cells were seeded at 50,000 cells per well in 24-well tissue culture plates. Cells were allowed to adhere overnight and were transfected for 24 h before media was replaced. A further 24 h after media replacement cells were treated with 1 μM 4-hydroxyestradiol (4-OHE2, Sigma) or 10 μM Mitomycin C (MMC, Sigma) for 3 h. Cells were gently washed with PBS and media was replaced with fresh complete media for a further 21 h before being assayed for DNA damage.

Immunocytochemistry. Cells were gently lifted from cell culture plates, cytospin onto slides, and fixed in ice-cold absolute methanol for 5 min. Slides were washed with PBS and blocked for 30 min with 1% bovine serum albumin in PBS-T (Tween-20 0.1% v/v). Slides were dual stained for 1 h with the mouse anti-phospho-H2AX (Ser139) antibody (1:500; ab26350, Abcam) and rabbit anti-53BP1 (1:500; ab36823, Abcam). Slides were incubated with anti-mouse AlexaFluor 488-conjugated (1:400; ab150113, Abcam) and anti-rabbit IgG-AlexaFluor 494-conjugated (1:400; ab150080, Abcam) secondary antibodies, and stained with DAPI for microscopic

examination. Images were taken at 40× magnification on the Axio Imager.Z1 Microscope (Zeiss). Co-localised γ-H2AX and 53BP1 foci were counted in >150 cells from a minimum of ten fields from three independent experiments.

Comet assay. Alkaline comet assays were performed using a comet assay kit (AbCam). Harvested cells were mixed with low melting agarose and transferred onto a glass slide covered in a base layer of agarose. Slides were immersed in lysis buffer for 60 min at 4 °C. Lysis buffer was replaced with alkaline solution (300 mM NaOH, pH 10, 1 mM EDTA) and samples were kept in the dark for 30 min. Slides were transferred to an electrophoresis chamber filled with alkaline solution and electrophoresis was performed for 20 min (1 V/cm). DNA was stained with Vista Green DNA Dye and images were captured by fluorescence microscopy on the Axio Imager.Z1 Microscope. Comets were scored using the CellProfiler software v3.1.8⁶⁹. Tail moments were assessed for >100 cells in three independent experiments.

Statistical analysis of in vitro data. All in vitro data were expressed as the mean ± standard error. The normality of data was visualised using the Q-Q plot and tested using the Shapiro-Wilk normality test. Statistical significance of differences between control and test groups were determined by an unpaired Student's *t*-test or two-way analysis of variance (ANOVA). All statistical tests were two sided and *p*-values <0.05 were considered significant.

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

Data availability

Genome-wide association summary statistics are available within the article. CIMBA phenotype data used in this study from BCFR-AU, BCFR-NC, BCFR-NY, BCFR-PA, BCFR-UT, BFOCC, BIDMC, MBMSA, CBCS, CNIO, COH, DEMOKRITOS, DFCL, FCCC, GEORGETOWN, HCSC, HRBCP, HUNBOCS, HVH, ICO, KCONFAB, KUMC, MAYO, MSKCC, MUV, NCI, NNPIO, NORTHSHORE, OSUCCG, PBCS, SMC, SWE-BRCA, UCHICAGO, UCSF, UPENN, UPITT, UTMADACC, VFCTG, and WCP studies are available in the dbGaP database under accession code phs001321.v1.p1. The complete dataset is not publicly available due to restraints imposed by the ethical committees of individual studies. Requests to access the complete dataset, which is subject to General Data Protection Regulation (GDPR) rules, can be made to the Data Access Coordinating Committee (DACC) of CIMBA, following the process described on the CIMBA website (<http://cimba.cege.medschl.cam.ac.uk/projects/data-access-requests/>). Submitted applications are reviewed by the CIMBA DACC every 3 months. CIMBA DACC approval is required to access data from studies BCFR-ON/OCGN, BFOCC-LV, BRICOH, CCGCRN, BRICOH, CONSIT TEAM, DKFZ, EMBRACE, FPGMX, GC-HBOC, GEMO, G-FAST, HEBCS, HEBON, IHCC, ILUH, INHERIT, IOVHBOCS, IPOBCS, KOHBRA, MCGILL, NCCS, NRG_ONCOLOGY, OUH, SEABASS, and UKGRFOCR (see Supplementary Data 12—for a list of all CIMBA studies). Summary statistics for each GWAS conducted for this study, can be freely downloaded from the NHGRI-EBI GWAS catalogue with the accession codes: GCST90134567; GCST90134568; GCST90134569; and GCST90134570; (<https://www.ebi.ac.uk/gwas/>). The source data for all figures are presented in the Supplementary Data file.

Code availability

Code for the retrospective likelihood analysis of variants is available on request.

Received: 15 November 2021; Accepted: 12 September 2022;
Published online: 06 October 2022

References

- Kuchenbaecker, K. B. et al. Risks of breast, ovarian, and contralateral breast cancer for *BRCA1* and *BRCA2* mutation carriers. *JAMA* **317**, 2402–2416 (2017).
- 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).
- Milne, R. L. & Antoniou, A. C. Modifiers of breast and ovarian cancer risks for *BRCA1* and *BRCA2* mutation carriers. *Endocr. Relat. Cancer* **23**, T69–T84 (2016).
- Coignard, J. et al. A case-only study to identify genetic modifiers of breast cancer risk for *BRCA1/BRCA2* mutation carriers. *Nat. Commun.* **12**, 1078 (2021).

5. Milne, R. L. et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat. Genet.* **49**, 1767–1778 (2017).
6. Zhang, H. et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* **52**, 572–581 (2020).
7. Barnes, D. R. et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet. Med.* **22**, 1653–1666 (2020).
8. Kuchenbaecker, K. B. et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J. Natl Cancer Inst.* **109**, djw302 (2017).
9. Abel, H. J. et al. Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **583**, 83–89 (2020).
10. Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **16**, 172–183 (2015).
11. Conrad, D. F. et al. Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
12. Kumaran, M. et al. Germline copy number variations are associated with breast cancer risk and prognosis. *Sci. Rep.* **7**, 1–15 (2017).
13. Long, J. et al. A common deletion in the APOBEC3 genes and breast cancer risk. *J. Natl Cancer Inst.* **105**, 573–579 (2013).
14. Walker, L. C. et al. Evaluation of copy-number variants as modifiers of breast and ovarian cancer risk for BRCA1 pathogenic variant carriers. *Eur. J. Hum. Genet.* **25**, 432–438 (2017).
15. Xuan, D. et al. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis* **34**, 2240–2243 (2013).
16. Walker, L. C. et al. Increased genomic burden of germline copy number variants is associated with early onset breast cancer: Australian breast cancer family registry. *Breast Cancer Res* **19**, 1–8 (2017).
17. Amos, C. I. et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol. Biomark. Prev.* **26**, 126–135 (2017).
18. Antoniou, A. C. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br. J. Cancer* **98**, 1457–1466 (2008).
19. Bojesen, S. E. et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* **45**, 371–384 (2013).
20. Savage, K. I. et al. BRCA1 deficiency exacerbates estrogen-induced DNA damage and genomic instability. *Cancer Res* **74**, 2773–2784 (2014).
21. Moynahan, M. E., Cui, T. Y. & Jasin, M. Homology-directed DNA repair, Mitomycin-C resistance, and chromosome stability is restored with correction of a Brca1 mutation. *Cancer Res* **61**, 4842–4850 (2001).
22. Bamber, D. E. et al. Phenol sulphotransferase SULT1A1*1 genotype is associated with reduced risk of colorectal cancer. *Pharmacogenetics* **11**, 679–685 (2001).
23. Liang, G., Miao, X., Zhou, Y., Tan, W. & Lin, D. A functional polymorphism in the SULT1A1 gene (G638A) is associated with risk of lung cancer in relation to tobacco smoking. *Carcinogenesis* **25**, 773–778 (2004).
24. Ozawa, S. et al. Association of genotypes of carcinogen-activating enzymes, phenol sulfotransferase SULT1A1 (ST1A3) and arylamine N-acetyltransferase NAT2, with urothelial cancer in a Japanese population. *Int. J. Cancer* **102**, 418–421 (2002).
25. Tang, D. et al. Sulfotransferase 1A1 (SULT1A1) polymorphism, PAH-DNA adduct levels in breast tissue and breast cancer risk in a case-control study. *Breast Cancer Res. Treat.* **78**, 217–222 (2003).
26. Kellen, E. et al. Does occupational exposure to PAHs, diesel and aromatic amines interact with smoking and metabolic genetic polymorphisms to increase the risk on bladder cancer? The Belgian case control study on bladder cancer risk. *Cancer Lett.* **245**, 51–60 (2007).
27. Khvastova, E. P., Pustynnyak, V. O. & Gulyaeva, L. F. Genetic polymorphism of estrogen metabolizing enzymes in Siberian women with breast cancer. *Genet. Test. Mol. Biomark.* **16**, 167–173 (2012).
28. Kotnis, A., Kannan, S., Sarin, R. & Mulherkar, R. Case-control study and meta-analysis of SULT1A1 Arg 213 His polymorphism for gene, ethnicity and environment interaction for cancer risk. *Br. J. Cancer* **99**, 1340–1347 (2008).
29. Lee, H. et al. SULT1A1 Arg213His polymorphism, smoked meat, and breast cancer risk: a case-control study and meta-analysis. *DNA Cell Biol.* **31**, 688–699 (2012).
30. Dumas, I. & Diorio, C. Estrogen pathway polymorphisms and mammographic density. *Anticancer Res* **31**, 4369–4386 (2011).
31. Gulyaeva, L. F. et al. Comparative analysis of SNP in estrogen-metabolizing enzymes for ovarian, endometrial, and breast cancers in Novosibirsk, Russia. *Adv. Exp. Med. Biol.* **617**, 359–366 (2008).
32. Syamala, V. S. et al. Possible risk modification by polymorphisms of estrogen metabolizing genes in familial breast cancer susceptibility in an Indian population. *Cancer Invest* **28**, 304–311 (2010).
33. Hebbing, S. J. et al. Human SULT1A1 gene: copy number differences and functional implications. *Hum. Mol. Genet.* **16**, 463–470 (2007).
34. Hempel, N., Gamage, N., Martin, J. L. & McManus, M. E. Human cytosolic sulfotransferase SULT1A1. *Int. J. Biochem. Cell Biol.* **39**, 685–689 (2007).
35. Ning, B. et al. Common genetic polymorphisms in the 5'-flanking region of the SULT1A1 gene: haplotypes and their association with platelet enzymatic activity. *Pharmacogenet. Genomics* **15**, 465–473 (2005).
36. Yu, X. et al. Copy number variation in sulfotransferase isoform 1A1 (SULT1A1) is significantly associated with enzymatic activity in Japanese subjects. *Pharmacogenomics Pers. Med.* **6**, 19–24 (2013).
37. Liu, J. et al. Relationship of SULT1A1 copy number variation with estrogen metabolism and human health. *J. Steroid Biochem. Mol. Biol.* **174**, 169–175 (2017).
38. Herrmann, K. et al. Formation of hepatic DNA adducts by methyleugenol in mouse models: drastic decrease by Sult1a1 knockout and strong increase by transgenic human SULT1A1/2. *Carcinogenesis* **35**, 935–941 (2014).
39. Sachse, B., Meinel, W., Glatt, H. & Monien, B. H. The effect of knockout of sulfotransferases 1a1 and 1d1 and of transgenic human sulfotransferases 1A1/1A2 on the formation of DNA adducts from furfuryl alcohol in mouse models. *Carcinogenesis* **35**, 2339–2345 (2014).
40. Tremmel, R. et al. Methyleugenol DNA adducts in human liver are associated with SULT1A1 copy number variations and expression levels. *Arch. Toxicol.* **91**, 3329–3339 (2017).
41. Dooley, T. P., Haldeman-Cahill, R., Joiner, J. & Wilborn, T. W. Expression profiling of human sulfotransferase and sulfatase gene superfamilies in epithelial tissues and cultured cells. *Biochem. Biophys. Res. Commun.* **277**, 236–245 (2000).
42. Falany, J. L. & Falany, C. N. Expression of cytosolic sulfotransferases in normal mammary epithelial cells and breast cancer cell lines. *Cancer Res* **56**, 1551–1555 (1996).
43. Rubin, G. L., Harrold, A. J., Mills, J. A., Falany, C. N. & Coughtrie, M. W. H. Regulation of sulphotransferase expression in the endometrium during the menstrual cycle, by oral contraceptives and during early pregnancy. *Mol. Hum. Reprod.* **5**, 995–1002 (1999).
44. Stanley, E. L., Hume, R., Visser, T. J. & Coughtrie, M. W. H. Differential expression of sulfotransferase enzymes involved in thyroid hormone metabolism during human placental development. *J. Clin. Endocrinol. Metab.* **86**, 5944–5955 (2001).
45. Mercer, K. E. et al. Expression of sulfotransferase isoform 1A1 (SULT1A1) in breast cancer cells significantly increases 4-hydroxytamoxifen-induced apoptosis. *Int. J. Mol. Epidemiol. Genet.* **1**, 92–103 (2010).
46. Harte, M. T. et al. BRD7, a subunit of SWI/SNF complexes, binds directly to BRCA1 and regulates BRCA1-dependent transcription. *Cancer Res* **70**, 2538–2547 (2010).
47. Adjei, A. A. & Weinsilboum, R. M. Catecholesterol sulfation: possible role in carcinogenesis. *Biochem. Biophys. Res. Commun.* **292**, 402–408 (2002).
48. Zhu, B. T. & Conney, A. H. Is 2-methoxyestradiol an endogenous estrogen metabolite that inhibits mammary carcinogenesis? *Cancer Res* **58**, 2269–2277 (1998).
49. Rebbeck, T. R. et al. Pairwise combinations of estrogen metabolism genotypes in postmenopausal breast cancer etiology. *Cancer Epidemiol. Biomark. Prev.* **16**, 444–450 (2007).
50. James, P. A. et al. Large genomic rearrangements in the familial breast and ovarian cancer gene BRCA1 are associated with an increased frequency of high risk features. *Fam. Cancer* **14**, 287–295 (2015).
51. Chen, S. & Parmigiani, G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J. Clin. Oncol.* **25**, 1329–1333 (2007).
52. Woodward, A. M., Davis, T. A., Silva, A. G. S., Kirk, J. A. & Leary, J. A. Large genomic rearrangements of both BRCA2 and BRCA1 are a feature of the inherited breast/ovarian cancer phenotype in selected families. *J. Med. Genet.* **42**, e31–e31 (2005).
53. Lovelock, P. K. et al. Identification of BRCA1 missense substitutions that confer partial functional activity: potential moderate risk variants? *Breast Cancer Res* **9**, R82 (2007).
54. Spurdle, A. B. et al. BRCA1 R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. *J. Med. Genet.* **49**, 525–532 (2012).
55. Rebbeck, T. R. et al. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. *JAMA* **313**, 1347–1361 (2015).
56. Carver, T. et al. CanRisk Tool—A web interface for the prediction of breast and ovarian cancer risk and the likelihood of carrying genetic pathogenic variants. *Cancer Epidemiol. Biomark. Prev.* **30**, 469–473 (2021).
57. Cooper, N. J. et al. Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes. *Hum. Mol. Genet* **24**, 1774–1790 (2015).
58. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).

59. Wang, K. et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**, 1665–1674 (2007).
60. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
61. Barnes, D. R. et al. Evaluation of association methods for analysing modifiers of disease risk in carriers of high-risk mutations. *Genet. Epidemiol.* **36**, 274–291 (2012).
62. Antoniou, A. C. et al. A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. *Genet. Epidemiol.* **29**, 1–11 (2005).
63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
64. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
65. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
66. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* **21**, 974–984 (2011).
67. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
68. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pii (2013).
69. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).

Acknowledgements

All the families and clinicians who contribute to the studies; Catherine M. Phelan for her contribution to CIMBA until she passed away on 22 September 2017; Sue Healey, in particular taking on the task of mutation classification with the late Olga Sinilnikova; Maggie Angelakos, Judi Maskiell, Gillian Dite, Helen Tsimiklis; members and participants in the New York site of the Breast Cancer Family Registry; members and participants in the Ontario Familial Breast Cancer Registry; Vilnius Rudaitis and Laimonas Griškevičius; Drs Janis Eglitis, Anna Krilova and Aivars Stengrevics; Yuan Chun Ding and Linda Steele for their work in participant enrolment and biospecimen and data management; Bent Ejlersen and Anne-Marie Gerdes for the recruitment and genetic counselling of participants; Alicia Barroso, Rosario Alonso and Guillermo Pita; all the individuals and the researchers who took part in CONSIT TEAM (Consorzio Italiano Tumori Ereditari Alla Mammella), in particular: Bernard Peissel, Dario Zimbalatti, Daniela Zaffaroni, Laura Ottini, Giuseppe Giannini and the personnel of the Cogentech Cancer Genetic Test Laboratory, Milan, Italy. Ms. JoEllen Weaver and Dr. Betsy Bove; FPGMX: members of the Cancer Genetics group (IDIS): Miguel Aguado, Olivia Fuentes, and Ana Crujeiras; IFE—Leipzig Research Centre for Civilization Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber); We thank all participants, clinicians, family doctors, researchers, and technicians for their contributions and commitment to the DKFZ study and the collaborating groups in Lahore, Pakistan (Muhammad U. Rashid, Noor Muhammad, Sidra Gull, Seerat Bajwa, Faiz Ali Khan, Humaira Naeemi, Saima Faisal, Asif Loya, Mohammed Aasim Yusuf) and Bogota, Colombia (Diana Torres, Ignacio Briceno, Fabian Gil). Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers (GEMO) study is a study from the National Cancer Genetics Network UNICANCER Genetic Group, France. We wish to pay a tribute to Olga M. Sinilnikova, who with Dominique Stoppa-Lyonnet initiated and coordinated GEMO until she sadly passed away on the 30th June 2014. The team in Lyon (Olga Sinilnikova, Mélanie Léoné, Laure Barjhoux, Carole Verny-Pierre, Sylvie Mazoyer, Francesca Damiola, Valérie Sorin) managed the GEMO samples until the biological resource centre was transferred to Paris in December 2015 (Noura Mebirouk, Fabienne Lesueur, Dominique Stoppa-Lyonnet). We want to thank all the GEMO collaborating groups for their contribution to this study: Coordinating Centre, Service de Génétique, Institut Curie, Paris, France: Muriel Belotti, Ophélie Bertrand, Anne-Marie Birot, Bruno Buecher, Sandrine Caputo, Chrystelle Colas, Anaïs Dupré, Emmanuelle Fourme, Marion Gauthier-Villars, Lisa Golmard, Marine Le Mentec, Virginie Moncoutier, Antoine de Pauw, Claire Saule, Dominique Stoppa-Lyonnet, and Inserm U900, Institut Curie, Paris, France: Fabienne Lesueur, Noura Mebirouk. Contributing Centres: Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon—Centre Léon Bérard, Lyon, France: Nadia Boutry-Kryza, Alain Calender, Sophie Giraud, Mélanie Léone. Institut Gustave Roussy, Villejuif, France: Brigitte Bressac-de-Paillerets, Olivier Caron, Marine Guillaud-Bataille. Centre Jean Perrin, Clermont–Ferrand, France: Yves-Jean Bignon, Nancy Uhrhammer. Centre Léon Bérard, Lyon, France: Valérie Bonadona, Christine Lasset. Centre François Bacless, Caen, France: Pascaline Berthet, Laurent Castera, Dominique Vaur. Institut Paoli Calmettes, Marseille, France: Violaine Bourdon, Catherine Nogués, Tetsuro Noguchi, Cornel Popovici, Audrey Remenieras, Hagay Sobol. CHU Arnaud-de-Villeneuve, Montpellier, France: Isabelle Coupier, Pascal Pujol. Centre Oscar Lambret, Lille, France: Claude Adenis, Aurélie Dumont, Françoise Révillon. Centre Paul Strauss, Strasbourg, France: Danièle Muller. Institut Bergonié, Bordeaux, France: Emmanuelle Barouk-Simonet, Françoise Bonnet, Virginie Bubié, Michel Longy, Nicolas Sevenet, Institut Claudius Regaud, Toulouse, France: Laurence Gladieff, Rosine

Guimbaud, Viviane Feillel, Christine Toulas. CHU Grenoble, France: Hélène Dreyfus, Christine Dominique Leroux, Magalie Peysselon, Rebischung. CHU Dijon, France: Amandine Baurand, Geoffrey Bertolone, Fanny Coron, Laurence Faivre, Caroline Jacquot, Sarah Lizard. CHU St-Etienne, France: Caroline Kientz, Marine Lebrun, Fabienne Prieur. Hôtel Dieu Centre Hospitalier, Chambéry, France: Sandra Fert Ferrer. Centre Antoine Lacassagne, Nice, France: Véronique Mari. CHU Limoges, France: Laurence Vénat-Bouvet. CHU Nantes, France: Stéphane Béziau, Capucine Delnatte. CHU Bretonneau, Tours and Centre Hospitalier de Bourges France: Isabelle Mortemousque. Groupe Hospitalier Pitié-Salpêtrière, Paris, France: Florence Coulet, Mathilde Warcoï. CHU Vandoeuvre-les-Nancy, France: Myriam Bronner, Johanna Sokolowska. CHU Besançon, France: Marie-Agnès Collonge-Rame, Alexandre Damette. CHU Poitiers, Centre Hospitalier d'Angoulême and Centre Hospitalier de Niort, France: Paul Gesta. Centre Hospitalier de La Rochelle: Hakima Lallaoui. CHU Nîmes Carêmeau, France: Jean Chiesa. CHI Poissy, France: Denise Molina-Gomes. CHU Angers, France: Olivier Ingster; CHU de Martinique, France: Odile Bera; Mickaëlle Rose; Ilse Coene and Brecht Crombez; Alicia Tosar and Paula Diaque; Virpi Palola; The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON) consists of the following Collaborating Centres: Netherlands Cancer Institute (coordinating centre), Amsterdam, NL: M.A. Rookus, F.B.L. Hogervorst, F.E. van Leeuwen, M.A. Adank, M.K. Schmidt, D.J. Jenner; Erasmus Medical Center, Rotterdam, NL: J.M. Collée, A.M.W. van den Ouweland, M.J. Hoening, I.A. Boere; Leiden University Medical Center, NL: C.J. van Asperen, P. Devilee, R.B. van der Luijt, T.C.T.E.F. van Cronenburg; Radboud University Nijmegen Medical Center, NL: M.R. Wevers, A.R. Mensenkamp; University Medical Center Utrecht, NL: M.G.E.M. Ausems, M.J. Koudijs; Amsterdam UMC, Univ of Amsterdam, NL: I. van de Beek; Amsterdam UMC, Vrije Universiteit Amsterdam, NL: K. van Engelen, J.J.P. Gille; Maastricht University Medical Center, NL: E.B. Gómez García, M.J. Blok, M. de Boer; University of Groningen, NL: L.P.V. Berger, A.H. van der Hout, M.J.E. Mourits, G.H. de Bock; The Netherlands Comprehensive Cancer Organisation (IKNL): S. Siesling, J. Verloop; The nationwide network and registry of histo- and cytopathology in The Netherlands (PALGA): E.C. van den Broek. HEBOS thanks Drs. Taru A. Muranen and Carl Blomqvist and research nurse Irja Erkkilä. HEBON thanks the study participants and the registration teams of IKNL and PALGA for part of the data collection; Hong Kong Sanatorium and Hospital; the Hungarian Breast and Ovarian Cancer Study Group members (Janos Papp, Aniko Bozsik, Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary) and the clinicians and patients for their contributions to this study; Hereditary Cancer Genetics Group, Vall d'Hebron Institute of Oncology (VHIO), the High Risk and Cancer Prevention Unit of the University Hospital Vall d'Hebron, and the Cellex Foundation for providing research facilities and equipment; the ICO Hereditary Cancer Program team led by Dr. Gabriel Capella; Dr Martine Dumont for sample management and skilful assistance; Catarina Santos and Pedro Pinto; members of the Center of Molecular Diagnosis, Oncogenetics Department and Molecular Oncology Research Center of Barretos Cancer Hospital; Heather Thorne, 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, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab; the KOBRA Study Group; Csilla Szabo (National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA); Lenka Foretova and Eva Machackova (Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute and MF MU, Brno, Czech Republic); and Michal Zikan, Petr Pohlreich and Zdenek Kleibl (Oncogynecologic Center and Department of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic); Anne Lincoln, Lauren Jacobs; the participants in Hereditary Breast/Ovarian Cancer Study and Breast Imaging Study for their selfless contributions to our research; the NISCC National Familial Cancer Consultation Service team led by Sara Dishon, the lab team led by Dr. Flavio Lejbkovic, and the research field operations team led by Dr. Mila Pinchev; the investigators of the Australia New Zealand NRG Oncology group; members and participants in the Ontario Cancer Genetics Network; Leigha Senter, Kevin Sweet, Caroline Craven, Julia Cooper, Amber Aielts, and Michelle O'Connor; Yip Cheng Har, Nur Aishah Mohd Taib, Phuah Sze Yee, Norhashimah Hassan and all the research nurses, research assistants and doctors involved in the MyBrCa Study for assistance in patient recruitment, data collection and sample preparation, Philip Iau, Sng Jen-Hwei and Sharifah Nor Akmal for contributing samples from the Singapore Breast Cancer Study and the HUKM-HKL Study respectively; the Meirav Comprehensive breast cancer center team at the Sheba Medical Center; Christina Selkirk; Håkan Olsson, Helena Jernström, Karin Henriksson, Katja Harbst, Maria Soller, Ulf Kristofferson; from Gothenburg Sahlgrenska University Hospital: Anna Öfverholm, Margareta Nordling, Per Karlsson, Zakaria Einbeigi; from Stockholm and Karolinska University Hospital: Anna von Wachenfeldt, Annelie Liljegren, Annika Lindblom, Brita Arver, Gisela Barbany Bustinza from Umeå University Hospital: Beatrice Melin, Christina Edwinstodter Ardnor, Monica Emanuelsson; from Uppsala University: Hans Ehrencrona, Maritta Hellström Pigg, Richard Rosenquist; from Linköping University Hospital: Marie Stenmark-Askmal, Sigrun Liedgren; Cecilia Zvocec, Qun Niu; Joyce Seldon and Lorna Kwan; Dr. Robert Nussbaum, Beth Crawford, Kate Loranger, Julie Mak, Nicola Stewart, Robin Lee, Amie Blanco and Peggy Conrad and Lalana Chan; Paul Pharoah, Carole Pye, Patricia Harrington and Eva Wozniak; Geoffrey Lindeman, Marion Harris, Martin Delatycki, Sarah Sawyer, Rebecca Driessen, and Ella Thompson for performing all DNA amplification. This study was supported by a grant from the New

Zealand Health Research Council (#17/113) and the Mackenzie Charitable Foundation. CIMBA: The CIMBA data management and data analysis were supported by Cancer Research—UK grants C12292/A20861 and PPRPGM-Nov20/100002 and the Gray Foundation. GCT and ABS are NHMRC Research Fellows. iCOGS: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer (CRN-87521), and 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. OncoArray: the PERSPECTIVE and PERSPECTIVE I&I projects funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec, and the Quebec Breast Cancer Foundation; the NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) initiative and Discovery, Biology and Risk of Inherited Variants in Breast Cancer (DRIVE) project (NIH Grants U19 CA148065 and X01HG007492); and Cancer Research UK (C1287/A10118 and C1287/A16563). BCFR: 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 centres 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. BFBOCC: Lithuania (BFBOCC-LT): Research Council of Lithuania grant SEN-18/2015. BIDMC: Breast Cancer Research Foundation. BMBSA: Cancer Association of South Africa (PI Elizabeth J. van Rensburg). CNIO study is partially funded by FIS PI19/00640 supported by FEDER funds, and the Spanish Network on Rare Diseases (CIBERER) CCGCRN: Research reported in this publication was supported by the Breast Cancer Research Foundation (project 20-172), National Cancer Institute of the National Institutes of Health under grant number R25CA112486, and RC4CA153828 (PI: J. Weitzel) from the National Cancer Institute and the Office of the Director, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. CONSIT TEAM: Associazione Italiana Ricerca sul Cancro (AIRC; IG2015 no.16732) to P. Peterlongo and P. Radice. 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/A23382 and C1287/A26886. D. Gareth Evans and Fiona Laloo are supported by an NIHR grant to the Biomedical Research Centre, Manchester (IS-BRC-1215-20007). The Investigators at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust are supported by an NIHR grant to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. Ros Eeles and Elizabeth Bancroft are supported by Cancer Research UK Grant C5047/A8385. Ros Eeles is also supported by NIHR support to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. FCCC: The University of Kansas Cancer Center (P30 CA168524) the Kansas Institute for Precision Medicine (P20GM130423) and the Kansas Bioscience Authority Eminent Scholar Program. A.K.G. was funded by R01CA140323, R01CA214545, R01CA260132, 5U10CA180888, and by the Chancellors Distinguished Chair in Biomedical Sciences Professorship. FPGMX: A.Vega is supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds (INT20/00071, INT17/00133, INT16/00154, INT15/00070; PI19/01424; PI16/00046; PI13/02030; PI10/00164), through the Autonomous Government of Galicia (Consolidation and structuring program: IN607B), through Centro de Investigación Biomédica en Red de Enfermedades Raras CIBERER (ACCI 2016: ER17P1AC7112/2018) and by the Fundación Mutua Madrileña (call 2018). GC-HBOC: German Cancer Aid (grant no 110837 and 113049, Rita K. Schmutzler) and the European Regional Development Fund and Free State of Saxony, Germany (LIFE—Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). GEMO: Ligue Nationale Contre le Cancer; the Association "Le cancer du sein, parlons-en!" Award, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program, the Fondation ARC pour la recherche sur le cancer (grant PJA 20151203365) and the French National Institute of Cancer (INCA grants AOR 01 082, 2001-2003, 2013-1-BCB-01-ICH-1 and SHS-E-SP 18-015). GEORGETOWN: the Survey, Recruitment, and Biospecimen Shared Resource at Georgetown University (NIH/NCI grant P30-CA051008) and the Fisher Center for Hereditary Cancer and Clinical Genomics Research. G-FAST: Bruce Poppe is a senior clinical investigator of FWO. Mattias Van Heetvelde obtained funding from IWT. HCSC: Spanish Ministry of Health PI15/00059, PI16/01292, and CB-161200301 CIBERONC from ISCIII (Spain), partially supported by European Regional Development FEDER funds. HEBCS: Helsinki University Hospital Research Fund, the Finnish Cancer Society and the Sigrid Juselius Foundation. HEBON: the Dutch Cancer Society grants NK11998-1854, NK12004-3088, NK12007-3756, the

Netherlands Organization of Scientific Research grant NWO 91109024, the Pink Ribbon grants 110005 and 2014-187.WO76, the BBMRI grant NWO 184.021.007/CP46 and the Transcan grant JTC 2012 Cancer 12-054. HEBON thanks the registration teams of Dutch Cancer Registry (IKNL; S. Siesling, J. Verloop) and the Dutch Pathology database (PALGA; L. Overbeek) for part of the data collection. HRBCP: Hong Kong Sanatorium and Hospital, Dr Ellen Li Charitable Foundation, The Kerry Group Kuok Foundation, National Institute of Health1R 03CA130065, and North California Cancer Center. HUNBOCS: Hungarian Research Grants KTIA-OTKA CK-80745 and NKFI_OTKA K-112228. (E. Olah) and 2019 Thematic Excellence Program (TUDFO/51757/2019-ITM) ICO: Supported by the Carlos III National Health Institute and Ministerio de Ciencia e Innovación, funded by FEDER funds—a way to build Europe—[PI19/00553; PI16/00563; PI18/01029 and CIBERONC]; the Government of Catalonia [Pla estratègic de recerca i innovació en salut (PERIS_MedPerCan and URDCat projects), 2017SGR1282 and 2017SGR496]. IHCC: The IHCC study was 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 financing 12,000,000 PLN. ILUH: Icelandic Association "Walking for Breast Cancer Research" and by the Landspítali University Hospital Research Fund. INHERIT: 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-SIIRI-701. IOVHBOCS: Ministero della Salute and "5 × 1000" Istituto Oncologico Veneto grant. IPOBCS: Liga Portuguesa Contra o Cancro. kConFab: The National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. KOHBRA: the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (H116C1127; 1020350; 1420190). KUMC: NIGMS P20 GM130423 (to AKG). MAYO: NIH grants CA116167, CA192393 and CA176785, an NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), and a grant from the Breast Cancer Research Foundation. MCGILL: Jewish 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 (2007Y2013)/European Research Council (Grant No. 310018). MODSQUAD: MH CZ—DRO (MMCI, 00209805), MEYS—NPS I—LO1413 to LF, and by Charles University in Prague project UNCE204024 (MZ). MSKCC: the Breast Cancer Research Foundation, the Robert and Kate Niehaus Clinical Cancer Genetics Initiative, the Andrew Sabin Research Fund and a Cancer Center Support Grant/Core Grant (P30 CA008748). NAROD: IR01 CA149429-01. NCI: the Intramural Research Program of the US National Cancer Institute, NIH, supports Mark H Greene and Megan Frone, and by support services contracts N02-CP-11019-50, N02-CP-21013-63 and N02-CP-65504 with Westat, Inc, Rockville, MD. NICCC: Clalit Health Services in Israel, the Israel Cancer Association and the Breast Cancer Research Foundation (BCRF), NY. NNPIO: the Russian Science Foundation (grant number 21-75-30015). 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 Tumors (ITT) grant 2014-2015-2016. SEABASS: Ministry of Science, Technology and Innovation, Ministry of Higher Education (UM.C/HIR/MOHE/06) and Cancer Research Initiatives Foundation. SMC: the Israeli Cancer Association. SWE-BRCA: the Swedish Cancer Society. UCHICAGO: NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA125183), R01 CA142996, 1U01CA161032 and by the Ralph and Marion Falk Medical Research Trust, the Entertainment Industry Fund National Women's Cancer Research Alliance and the Breast Cancer research Foundation. O.I.O. is an ACS Clinical Research Professor. UCLA: Jonsson Comprehensive Cancer Center Foundation; Breast Cancer Research Foundation. UCSF: UCSF Cancer Risk Program and Helen Diller Family Comprehensive Cancer Center. UKFOCR: Cancer Research UK. UPENN: Breast Cancer Research Foundation; Susan G. Komen Foundation for the cure, Bassett Research Center for BRCA. UPITT/MWH: Hackers for Hope Pittsburgh. VFCTG: Victorian Cancer Agency, Cancer Australia, National Breast Cancer Foundation. WCP: 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 UL1TR000124. L.C.W. was funded by the Rutherford Discovery Fellowship (Royal Society of New Zealand)—RDF-U001501. A.B.S. was supported by an NHMRC Investigator Fellowship (APP177524). T.N.-D. is a National Breast Cancer Foundation (Australia) Career Development Fellow. J.D. is supported by the CanRisk Cancer Research UK programme grant: PPRPGM-Nov20/100002 and by the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health. MT was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014).

Author contributions

L.C.W., A.C.A., A.B.S., J.F.P., B.A.R., and P.D.M. conceived the study design. C.H. and L.C.W. drafted the initial manuscript, while the complete writing group consisted of

C.H., L.C.W., A.C.A., A.B.S., L.Ma., J.D. G.C.-T., J.F.P., B.A.R., P.D.M., and D.R.B. L.Ma. and D.R.B. performed the statistical analyses. J.D. conducted the CNV calling. C.H. conducted the laboratory experiments and G.A.R.W. generated the genetically modified cell line. J.D., L.Mc., and G.Le. performed the data management. K.Ai., I.L.A., B.K.A., J.A., J.Ba., R.B.B., S.B., L.B., M.J.B., S.E.B., J.Bo., A.R.B., J.Br., S.S.B., M.A.C., I.C., W.K.C., K.B.M.C., GEMO Study Collaborators, EMBRACE Collaborators, M.-A.C.-R., J.Co., C.C., F.J.C., M.B.D., S.D., R.Da., M.d.l.h., R.d.P., C.D., M.D., O.D., Y.C.D., S.M.D., A.Do., J.E., D.F.E., H.E., C.E., D.G.E., U.F., L.F., F.F., E.Fr., M.F., D.F., J.G., S.A.G., A.G., P.Ge., A.K.G., D.E.G., M.H.G., E.Ha., C.H., U.H., T.V.O.H., J.Ha., J.He., N.He., E.Hon., P.J.H., E.N.I., SWE-BRCA Investigators, kConFab Investigators, HEBON Investigators, C.I., L.I., A.I., A.Ja., P.A.J., R.J., E.M.J., V.J., B.Y.K., Z.K., J.K., I.K., M.K., A.K., Y.L., F.L., C.Las., C.Lau., C.Laz., C.Le, F.Les., P.L.M., S.M., V.M., J.W.M.M., N.M., A.Me., A.Mil., M.M., L.Mo., E.M.-F., H.M., S.N., K.L.N., S.L.N., H.N., J.N.Y.Y., T.N.-D., L.N.-Z., K.O., E.O., O.L.O., A.O., C.-E.O., S.K.P., M.T.P., I.S.P., A.P., P.P.-S., P.P., T.P., P.R., J.Ram., J.Ran., G.C.R., K.R., E.H.R., M.R., R.K.S., P.D.S., S.S., P.S., L.E.S., J.Si., C.F.S., K.S., D.S., D.S.-L., C.S., Y.Y.T., M.R.T., S.H.T., M.Th., D.L.T., M.Ti., A.E.T., A.H.T., V.T., N.T., K.v.E., E.J.v.R., A.Ve., A.Vi., L.W., J.N.W., and M.R.W. provided DNA samples and 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/s42003-022-03978-6>.

Correspondence and requests for materials should be addressed to Logan C. Walker.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: H el ene Choquet and Eve Rogers. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>




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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

  The Author(s) 2022

Christopher Hakkaart¹, John F. Pearson ¹, Louise Marquart^{2,3}, Joe Dennis ⁴, George A. R. Wiggins¹, Daniel R. Barnes ⁴, Bridget A. Robinson^{5,6}, Peter D. Mace ⁷, Kristiina Aittom aki⁸, Irene L. Andrulis ^{9,10}, Banu K. Arun¹¹, Jacopo Azzollini ¹², Judith Balma a^{13,14}, Rosa B. Barkardottir^{15,16}, Sami Belhadj¹⁷, Lieke Berger¹⁸, Marinus J. Blok ¹⁹, Susanne E. Boonen²⁰, Julika Borde^{21,22,23}, Angela R. Bradbury²⁴, Joan Brunet ²⁵, Saundra S. Buys²⁶, Maria A. Caligo²⁷, Ian Campbell ^{28,29}, Wendy K. Chung ³⁰, Kathleen B. M. Claes ³¹, GEMO Study Collaborators*, EMBRACE Collaborators*, Marie-Agn es Collonge-Rame³², Jackie Cook³³, Casey Cosgrove³⁴, Fergus J. Couch³⁵, Mary B. Daly³⁶, Sita Dandiker ¹⁷, Rosemarie Davidson³⁷, Miguel de la Hoya ³⁸, Robin de Putter ³¹, Capucine Delnatte³⁹, Mallika Dhawan⁴⁰, Orland Diez^{13,41}, Yuan Chun Ding⁴², Susan M. Domchek⁴³, Alan Donaldson⁴⁴, Jacqueline Eason⁴⁵, Douglas F. Easton ^{4,46}, Hans Ehrencrona ^{47,48}, Christoph Engel^{49,50}, D. Gareth Evans^{51,52}, Ulrike Faust⁵³, Lidia Feliubadal o ²⁵, Florentia Fostira⁵⁴, Eitan Friedman^{55,56}, Megan Frone⁵⁷, Debra Frost⁴, Judy Garber⁵⁸, Simon A. Gayther ⁵⁹, Andrea Gehrig⁶⁰, Paul Gesta ⁶¹, Andrew K. Godwin ⁶², David E. Goldgar⁶³, Mark H. Greene ⁵⁷, Eric Hahnen^{21,23}, Christopher R. Hake⁶⁴, Ute Hamann⁶⁵, Thomas V. O. Hansen⁶⁶, Jan Hauke ^{21,22,23}, Julia Hentschel⁶⁷, Natalie Herold^{21,22,23}, Ellen Honisch⁶⁸, Peter J. Hulick ^{69,70}, Evgeny N. Imyanitov⁷¹, SWE-BRCA Investigators*, kConFab Investigators*, HEBON Investigators*, Claudine Isaacs ⁷², Louise Izatt ⁷³, Angel Izquierdo ²⁵, Anna Jakubowska ^{74,75}, Paul A. James ^{29,76}, Ramunas Janavicius^{77,78}, Esther M. John^{79,80}, Vijai Joseph ¹⁷, Beth Y. Karlan ⁸¹, Zoe Kemp⁸², Judy Kirk⁸³, Irene Konstantopoulou ⁵⁴, Marco Koudijs⁸⁴, Ava Kwong^{85,86,87}, Yael Laitman⁵⁵, Fiona Lalloo⁸⁸, Christine Lasset⁸⁹, Charlotte Lautrup⁹⁰, Conxi Lazaro ²⁵, Cl ementine Legrand⁹¹, Goska Leslie ⁴, Fabienne Lesueur ^{92,93,94}, Phuong L. Mai⁹⁵, Siranoush Manoukian¹², V eronique Mari⁹⁶, John W. M. Martens ⁹⁷, Lesley McGuffog⁴, Noura Mebirouk^{92,93,94}, Alfons Meindl⁹⁸, Austin Miller ⁹⁹, Marco Montagna¹⁰⁰, Lidia Moserle¹⁰⁰, Emmanuelle Mouret-Fourme¹⁰¹, Hannah Musgrave¹⁰², Sophie Nambot ¹⁰³, Katherine L. Nathanson ⁴³, Susan L. Neuhausen ⁴², Heli Nevanlinna ¹⁰⁴, Joanne Ngeow Yuen Yie ^{105,106}, Tu Nguyen-Dumont^{107,108}, Liene Nikitina-Zake ¹⁰⁹, Kenneth Offit^{17,110}, Edith Olah¹¹¹, Olufunmilayo I. Olopade ¹¹², Ana Osorio ¹¹³, Claus-Eric Ott ¹¹⁴, Sue K. Park ^{115,116,117},

Michael T. Parsons ¹¹⁸, Inge Sokilde Pedersen ^{119,120,121}, Ana Peixoto¹²², Pedro Perez-Segura³⁸, Paolo Peterlongo ¹²³, Timea Pocza¹¹¹, Paolo Radice ¹²⁴, Juliane Ramser¹²⁵, Johanna Rantala¹²⁶, Gustavo C. Rodriguez¹²⁷, Karina Rønlund¹²⁸, Efraim H. Rosenberg ¹²⁹, Maria Rossing^{130,131}, Rita K. Schmutzler^{21,22,23}, Payal D. Shah²⁴, Saba Sharif¹³², Priyanka Sharma¹³³, Lucy E. Side¹³⁴, Jacques Simard ¹³⁵, Christian F. Singer¹³⁶, Katie Snape¹³⁷, Doris Steinemann ¹³⁸, Dominique Stoppa-Lyonnet^{101,139,140}, Christian Sutter ¹⁴¹, Yen Yen Tan ¹³⁶, Manuel R. Teixeira ^{122,142}, Soo Hwang Teo ^{143,144}, Mads Thomassen²⁰, Darcy L. Thull¹⁴⁵, Marc Tischkowitz ^{146,147}, Amanda E. Toland ¹⁴⁸, Alison H. Trainer^{76,149}, Vishakha Tripathi ¹⁵⁰, Nadine Tung¹⁵¹, Klaartje van Engelen¹⁵², Elizabeth J. van Rensburg ¹⁵³, Ana Vega ^{154,155,156}, Alessandra Viel ¹⁵⁷, Lisa Walker¹⁵⁸, Jeffrey N. Weitzel ¹⁵⁹, Marike R. Wevers ¹⁶⁰, Georgia Chenevix-Trench¹¹⁸, Amanda B. Spurdle ¹¹⁸, Antonis C. Antoniou⁴ & Logan C. Walker ^{1✉}

¹Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand. ²QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ³School of Public Health, University of Queensland, Brisbane, Australia. ⁴Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁵Department of Medicine, University of Otago, Christchurch, New Zealand. ⁶Canterbury Regional Cancer and Haematology Service, Canterbury District Health Board, Christchurch Hospital, Christchurch, New Zealand. ⁷Department of Biochemistry, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand. ⁸Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland. ⁹Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada. ¹⁰Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada. ¹¹Department of Breast Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹²Unit of Medical Genetics, Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy. ¹³Hereditary cancer Genetics Group, Vall d'Hebron Institute of Oncology, Vall d'Hebron Hospital Campus, Barcelona, Spain. ¹⁴Department of Medical Oncology, Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. ¹⁵Department of Pathology, Landspítali University Hospital, Reykjavík, Iceland. ¹⁶BMC (Biomedical Centre), Faculty of Medicine, University of Iceland, Reykjavík, Iceland. ¹⁷Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹⁸Department of Clinical Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. ¹⁹Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands. ²⁰Department of Clinical Genetics, Odense University Hospital, Odense C, Denmark. ²¹Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ²²Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ²³Center for Familial Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. ²⁴Department of Medicine, Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. ²⁵Hereditary Cancer Program, Catalan Institute of Oncology (ICO), ONCOBELL-IDIBELL-IGTP, CIBERONC, Barcelona, Spain. ²⁶Department of Medicine, Huntsman Cancer Institute, Salt Lake City, UT, USA. ²⁷SOD Genetica Molecolare, University Hospital, Pisa, Italy. ²⁸Peter MacCallum Cancer Center, Melbourne, Victoria, Australia. ²⁹Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Victoria, Australia. ³⁰Departments of Pediatrics and Medicine, Columbia University, New York, NY, USA. ³¹Centre for Medical Genetics, Ghent University Hospital, Ghent, Belgium. ³²Service de Génétique Biologique, CHRU de Besançon, Besançon, France. ³³Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Sheffield, UK. ³⁴Gynecologic Oncology, Translational Therapeutics, Department of Obstetrics and Gynecology, Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. ³⁵Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ³⁶Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA. ³⁷Department of Clinical Genetics, Queen Elizabeth University Hospital, Glasgow, UK. ³⁸Molecular Oncology Laboratory, CIBERONC, Hospital Clínico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain. ³⁹Oncogénétique, Institut de Cancérologie de l'Ouest site René Gauducheau, Saint Herblain, France. ⁴⁰Cancer Genetics and Prevention Program, University of California San Francisco, San Francisco, CA, USA. ⁴¹Area of Clinical and Molecular Genetics, Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. ⁴²Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA. ⁴³Basser Center for BRCA, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA. ⁴⁴Clinical Genetics Department, St Michael's Hospital, Bristol, UK. ⁴⁵Nottingham Clinical Genetics Service, Nottingham University Hospitals NHS Trust, Nottingham, UK. ⁴⁶Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. ⁴⁷Department of Clinical Genetics and Pathology, Laboratory Medicine, Skåne University Hospital, Lund, Sweden. ⁴⁸Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, Lund, Sweden. ⁴⁹Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. ⁵⁰LIFE - Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany. ⁵¹Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. ⁵²North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. ⁵³Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany. ⁵⁴Molecular Diagnostics Laboratory, INRASTES, National Centre for Scientific Research 'Demokritos', Athens, Greece. ⁵⁵The Susanne Levy Gertner Oncogenetics Unit, Chaim Sheba Medical Center, Ramat Gan, Israel. ⁵⁶Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel. ⁵⁷Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. ⁵⁸Cancer Risk and Prevention Clinic, Dana-Farber Cancer Institute, Boston, MA, USA. ⁵⁹Center for Bioinformatics and Functional Genomics and the Cedars Sinai Genomics Core, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ⁶⁰Department of Human Genetics, University Würzburg, Würzburg, Germany. ⁶¹Service Régional Oncogénétique Poitou-Charentes, CH Niort, Niort, France. ⁶²Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS, USA. ⁶³Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA. ⁶⁴Waukesha Memorial Hospital-Pro Health Care, Waukesha, USA. ⁶⁵Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁶⁶Department of Clinical Genetics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark.

⁶⁷Institute of Human Genetics, University Hospital Leipzig, Leipzig, Germany. ⁶⁸Department of Gynecology and Obstetrics, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany. ⁶⁹Center for Medical Genetics, NorthShore University HealthSystem, Evanston, IL, USA. ⁷⁰The University of Chicago Pritzker School of Medicine, Chicago, IL, USA. ⁷¹N.N. Petrov Institute of Oncology, St. Petersburg, Russia. ⁷²Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA. ⁷³Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, London, UK. ⁷⁴Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. ⁷⁵Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland. ⁷⁶Parkville Familial Cancer Centre, Peter MacCallum Cancer Center, Melbourne, Victoria, Australia. ⁷⁷Faculty of Medicine, Institute of Biomedical Sciences, Dept. Of Human and Medical Genetics, Vilnius University, Vilnius, Lithuania. ⁷⁸State Research Institute Centre for Innovative Medicine, Vilnius, Lithuania. ⁷⁹Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA. ⁸⁰Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. ⁸¹David Geffen School of Medicine, Department of Obstetrics and Gynecology, University of California at Los Angeles, Los Angeles, CA, USA. ⁸²Breast and Cancer Genetics Units, The Royal Marsden NHS Foundation Trust, London, UK. ⁸³Familial Cancer Service, Weatmead Hospital, Wentworthville, New South Wales, Australia. ⁸⁴Department of Medical Genetics, University Medical Center, Utrecht, The Netherlands. ⁸⁵Hong Kong Hereditary Breast Cancer Family Registry, Hong Kong, China. ⁸⁶Department of Surgery, The University of Hong Kong, Hong Kong, China. ⁸⁷Department of Surgery and Cancer Genetics Center, Hong Kong Sanatorium and Hospital, Hong Kong, China. ⁸⁸Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA. ⁸⁹Unité de Prévention et d'Epidémiologie Génétique, Centre Léon Bérard, Lyon, France. ⁹⁰Department of Clinical Genetics, Aarhus University Hospital, Aarhus N, Denmark. ⁹¹Département de Génétique, CHU de Grenoble, Grenoble, France. ⁹²Genetic Epidemiology of Cancer team, Inserm U900, Paris, France. ⁹³Institut Curie, Paris, France. ⁹⁴Mines ParisTech, Fontainebleau, France. ⁹⁵Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ⁹⁶Département d'Hématologie-Oncologie Médicale, Centre Antoine Lacassagne, Nice, France. ⁹⁷Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands. ⁹⁸Department of Gynecology and Obstetrics, University of Munich, Campus Großhadern, Munich, Germany. ⁹⁹NRG Oncology, Statistics and Data Management Center, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA. ¹⁰⁰Immunology and Molecular Oncology Unit, Veneto Institute of Oncology IOV - IRCCS, Padua, Italy. ¹⁰¹Service de Génétique, Institut Curie, Paris, France. ¹⁰²Department of Clinical Genetics, Yorkshire Regional Genetics Service, Chapel Allerton Hospital, Leeds, UK. ¹⁰³Unité d'oncogénétique, Centre de Lutte Contre le Cancer, Centre Georges-François Leclerc, Dijon, France. ¹⁰⁴Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. ¹⁰⁵Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. ¹⁰⁶Cancer Genetics Service, National Cancer Centre, Singapore, Singapore. ¹⁰⁷Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia. ¹⁰⁸Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia. ¹⁰⁹Latvian Biomedical Research and Study Centre, Riga, Latvia. ¹¹⁰Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹¹¹Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary. ¹¹²Center for Clinical Cancer Genetics, The University of Chicago, Chicago, IL, USA. ¹¹³Familial Cancer Clinical Unit, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO) and Spanish Network on Rare Diseases (CIBERER), Madrid, Spain. ¹¹⁴Institute of Medical Genetics and Human Genetics, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Berlin, Germany. ¹¹⁵Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. ¹¹⁶Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, South Korea. ¹¹⁷Cancer Research Institute, Seoul National University, Seoul, Korea. ¹¹⁸Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. ¹¹⁹Molecular Diagnostics, Aalborg University Hospital, Aalborg, Denmark. ¹²⁰Clinical Cancer Research Center, Aalborg University Hospital, Aalborg, Denmark. ¹²¹Department of Clinical Medicine, Aalborg University, Aalborg, Denmark. ¹²²Department of Genetics, Portuguese Oncology Institute, Porto, Portugal. ¹²³Genome Diagnostics Program, IFOM ETS - the AIRC Institute of Molecular Oncology, Milan, Italy. ¹²⁴Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy. ¹²⁵Division of Gynaecology and Obstetrics, Klinikum rechts der Isar der Technischen Universität München, Munich, Germany. ¹²⁶Clinical Genetics, Karolinska Institutet, Stockholm, Sweden. ¹²⁷Division of Gynecologic Oncology, NorthShore University HealthSystem, University of Chicago, Evanston, IL, USA. ¹²⁸Department of Clinical Genetics, University Hospital of Southern Denmark, Vejle Hospital, Vejle, Denmark. ¹²⁹Department of Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. ¹³⁰Center for Genomic Medicine, Rigshospitalet, Copenhagen, Denmark. ¹³¹Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. ¹³²West Midlands Regional Genetics Service, Birmingham Women's Hospital Healthcare NHS Trust, Birmingham, UK. ¹³³Department of Internal Medicine, Division of Medical Oncology, University of Kansas Medical Center, Westwood, KS, USA. ¹³⁴Princess Anne Hospital, Southampton, UK. ¹³⁵Genomics Center, Centre Hospitalier Universitaire de Québec - Université Laval Research Center, Québec City, QC, Canada. ¹³⁶Dept of OB/GYN and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria. ¹³⁷Medical Genetics Unit, St George's, University of London, London, UK. ¹³⁸Institute of Human Genetics, Hannover Medical School, Hannover, Germany. ¹³⁹Department of Tumour Biology, INSERM U830, Paris, France. ¹⁴⁰Université Paris Cité, Paris, France. ¹⁴¹Institute of Human Genetics, University Hospital Heidelberg, Heidelberg, Germany. ¹⁴²Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal. ¹⁴³Breast Cancer Research Programme, Cancer Research Malaysia, Subang Jaya, Selangor, Malaysia. ¹⁴⁴Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. ¹⁴⁵Department of Medicine, Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ¹⁴⁶Program in Cancer Genetics, Departments of Human Genetics and Oncology, McGill University, Montréal, QC, Canada. ¹⁴⁷Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK. ¹⁴⁸Department of Cancer Biology and Genetics, The Ohio State University, Columbus, OH, USA. ¹⁴⁹Department of medicine, University Of Melbourne, Melbourne, Victoria, Australia. ¹⁵⁰South East Thames Regional Genetics Service, Guy's and St Thomas' NHS Foundation Trust, London, UK. ¹⁵¹Department of Medical Oncology, Beth Israel Deaconess Medical Center, Boston, MA, USA. ¹⁵²Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands. ¹⁵³Department of Genetics, University of Pretoria, Arcadia, South Africa. ¹⁵⁴Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain. ¹⁵⁵Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain. ¹⁵⁶Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain. ¹⁵⁷Division of Functional onco-genomics and genetics, Centro di Riferimento Oncologico di Aviano (CRO), IRCCS, Aviano, Italy. ¹⁵⁸Oxford Regional Genetics Service, Churchill Hospital, Oxford, UK. ¹⁵⁹Latin American School of Oncology, Tuxtla Gutiérrez, Chiapas, Mexico. ¹⁶⁰Radboud University Medical Center, Nijmegen, Netherlands. *Lists of authors and their affiliations appear at the end of the paper. ✉email: logan.walker@otago.ac.nz

GEMO Study Collaborators

Marie-Agnès Collonge-Rame³², Capucine Delnatte³⁹, Paul Gesta⁶¹, Christine Lasset⁸⁹, Clémentine Legrand⁹¹, Fabienne Lesueur^{92,93,94}, Véronique Mari⁹⁶, Noura Mebirouk^{92,93,94}, Emmanuelle Mouret-Fourme¹⁰¹, Sophie Nambot¹⁰³ & Dominique Stoppa-Lyonnet^{101,139,140}

EMBRACE Collaborators

Jackie Cook³³, Rosemarie Davidson³⁷, Alan Donaldson⁴⁴, Jacqueline Eason⁴⁵, D. Gareth Evans^{51,52}, Debra Frost⁴, Louise Izatt⁷³, Zoe Kemp⁸², Fiona Lalloo⁸⁸, Hannah Musgrave¹⁰², Saba Sharif¹³², Lucy E. Side¹³⁴, Katie Snape¹³⁷, Vishakha Tripathi¹⁵⁰ & Lisa Walker¹⁵⁸

SWE-BRCA Investigators

Hans Ehrencrona^{47,48} & Johanna Rantala¹²⁶

kConFab Investigators

Georgia Chenevix-Trench¹¹⁸ & Amanda B. Spurdle¹¹⁸

HEBON Investigators

Lieke Berger¹⁸, Marinus J. Blok¹⁹, Marco Koudijs⁸⁴, John W. M. Martens⁹⁷, Efraim H. Rosenberg¹²⁹, Klaartje van Engelen¹⁵² & Marijke R. Wevers¹⁶⁰