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 carriers. 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 carriers.\(^9\)

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 regulation.\(^11\) Germ line 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 risk.\(^14\) Although no variant reached the widely-adopted genome-wide statistical significance threshold applied for SNP-centred 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 suggest 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 Network.\(^17\) 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 called CNVs overlapped were the same variants. For example, 54.1% of BRCA1 deletions not detected by PennCNV were the same variant (c.5333_5346 + 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) CNVs called 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 map.\(^10\) 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.

### Prioritisation 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 map\(^10\) 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—\(p \leq 1 \times 10^{-3}\); 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
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 × 10^{-3}) and BRCA2 carriers (HR = 0.44, 95% CI 0.22–0.88, p = 9.2 × 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, suggested increased risk (HR = 1.92, 95% CI = 1.06–3.46, p = 6.0 × 10^{-3}) and decreased risk (HR = 0.13, 95% CI = 0.04–0.45, p = 3.3 × 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 CNVs.

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 × 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–WT) 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 × 10^{-4}) reduction in ESR1 expression (Fig. 2d, Supplementary Data 10) and a significant 78% (p = 2.03 × 10^{-3}) increase in CYP1A1 expression (Fig. 2e, Supplementary Data 10), consistent with breast cells with a pathogenic BRCA1 variant.

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+/−

| Table 1 Breast cancer hazard ratio estimates using a single model comparing other BRCA1/2 pathogenic variants versus (1) deletions, and (2) duplications. |
| --- |
| | Mutation type | Unaffected | Affected | HR | 95% CI | p |
| BRCA1 carriers | Other | 8669 | 8483 | 1.00 | [reference] |
| | Deletions | 636 | 789 | 1.21 | 1.09 | 1.35 | 4.35E-04 |
| | Duplications | 153 | 192 | 0.99 | 1.48 | 6.60E-02 |
| BRCA2 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 |

* Weighted cohort models fitted separately for BRCA1 and BRCA2 carriers. Weights were calculated assuming BRCA1/2 carrier breast cancer incidence 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 | Proportion validated |
| SULT1A1 | Deletion | BRCA1 | 1 | Yes | 100% (8/8) |
| TERT | Deletion | BRCA2 | 1 | No | 0% (0/1) |
| | | | 2 | No | 0% (0/1) |
| | | | 3 | No | 33% (1/3) |
| LSP1 | Duplication | BRCA2 | 1 | No | 50% (1/2) |

CNV copy number variant. Zarrei et al (2015) stringent CNV map.
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 × 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 γH2AX and 53BP1 immunostaining assays. Analysis using comet assays, showed that treatment with 4-OHE2 (F(1, 8) = 5.79, p = 4.28 × 10^{-2}; Fig. 3e, Supplementary Data 11) and MMC (F(1, 8) = 5.73, p = 4.36 × 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 × 10^{-2}; Fig. 3e, Supplementary Data 11) and MMC (F(1, 8) = 5.69, p = 4.43 × 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 × 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
reach statistical significance ($F(1, 8) = 5.02, p = 5.5 \times 10^{-2}$). Additionally, for the MMC treated MCF7–BRCA1$^{1/−}$ cells there was a significant interaction between the siSULT1A1 transfection and MMC treatment ($F(1, 4) = 8.44, p = 4.4 \times 10^{-2}$) 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/53BP1 foci 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 map\(^{(10)}\). 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, \( p = 4.3 \times 10^{-3} \))\(^{32}\), 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 phenotype\(^{37}\). This finding corresponds with results from phenotypic analyses of mouse SULT1A1 knockout mice 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 activity\(^{40}\). 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\(^{30,46}\). However, levels of CYP1A1 are higher in breast cells lacking BRCA1 function and promote the formation of the carcinogenic 2-hydroxyestradiol (2-OHE2)\(^{30}\). 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\(^{37–39}\) 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 cancer\(^{48}\). 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*2C) that has increased inducibility to produce catechol oestrogens\(^{49}\). For European-Americans, carrying the CYP1A1*2C genotype was associated with increased breast cancer risk (odds ratio (OR) = 1.71, 95% CI = 1.09–2.67). However, carrying the CYP1A1*2C 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*2C allele. There was no association between the CYP1A1*2C (rs1048943) polymorphism and breast cancer risk in BRCA1 pathogenic variant carriers\(^{5}\). These results further suggest that the balance between the generation of catecholes-trogens and catechol estrogen 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 Indels\(^{50}\). 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\(^{33,34}\). 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\(^{1,35}\). 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://degc.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 an increase in risk of breast cancer in BRCA1 pathogenic variant carriers. If verified, future therapeutic interventions directly 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/) along with family history, lifestyle/hormonal risk factors, common genetic susceptibility variants, and marmomorphologic 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 participants across the United States, California, Europe, and Australia participating in CMBA (Supplementary Data 12), as reported previously43. 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 non-redundant gene-centric CNVs 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 quality control measures were performed using the Illumina GenomeStudio (Illumina) and QC software to assess the quality of the 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)37 in the statistical program R (V3.5.3)38. CNV calls were generated using PennCNV39. 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 elsewhere9. A total of 13679 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, RAF std > 0.01, waviness factor deviating from 0 by ≥ 0.05; LRR outliers < 1.0, BAF s.d. ≥ 0.2, LRR s.d. ≥ 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) Genome Browser tracks. A total of 30,336 gene regions with 27,038 unique gene symbols were derived and used in our research. These gene regions by one or more consistent deletions were identified in a genome-wide scan in R (V3.5.3) using the GenomicRanges package (V1.4)40. 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 up to the age of 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 ascertainment44. This model is unable to adjust for covariates. An approximation method yields HR and 95%CI estimates based on this score test41. 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 recalled for the analysis of overlapping gene regions after quality control). The association 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 ≤ 0.05/0.05 = 8 × 10⁻⁵; duplications in BRCA1 carriers—p ≤ 0.05/0.05 = 10⁻⁶; deletions in BRCA2 carriers—p ≤ 0.05/0.05 = 10⁻⁶; and duplications in BRCA2 carriers—p ≤ 0.05/0.8469 = 6 × 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 list.

Models to estimate the associations (HRs) of deletions and duplications simultaneously took the form of a weighted cohort analysis45,46. 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 approach developed by 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, ≥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 assessed by MLPA. Customised MLPA probes and probe sequences were derived from life-time CNV validation of Life Technologies used to validate CNVs are listed in Supplementary Data 15. SULT1A1 MLPA was performed using the SULTL1A1 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 × 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 v.3.7.20 best practice guidelines. Cases were sequenced to an average of 30-fold depth and CNVs were called using Lumpy and CNVnator (Supplementary Data 14). 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 (v7.2.7a-78+95f), followed by Novosort (V1.03.01) 01 create coordinate-sorted duplicate marked files. GATK (V3.3) Indel realignment and base quality recalibration were used to create analysis ready reads49. 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 (V0.213) and read depth differences using CNVnator (V0.3.3).

**Dosage effect analysis.** Expression and copy number data from the Breast Invasive Carcinoma datasets were downloaded using cBioPortal. 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 passed every 3–4 days. Cells were used up to a maximum of 30 passages.

**Development of MCF7–BRCA1Δ11 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 termed a pathogenic variant). The guide RNA was designed to target exon 11 of BRCA1 and disrupt its function (sequence 5'-CGACGACTTGGAAACCCCA-3'). MCF7 cells were transfected with plasmid containing gRNA, Cas9 protein, and puromycin resistance (Addgene [ID #62988]–pSPCas9(BB)-2A-Puro (PX595 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 resistance (Addgene [ID #62988]) plasmid was Sanger sequenced to identify cleavage site was amplification of the Cas9 locus.

**Cell culture.** 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 passed every 3–4 days. Cells were used up to a maximum of 30 passages.

**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 BRCA1 (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 the 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 RNAgenn-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. Genespecific 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–BRCA1Δ11 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 cTM 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 transfected 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 24 h before being assayed for DNA damage.

**Immunocytochemistry.** Cells were gently lifted from cell culture plates, cytospun 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 then stained for 1 h with the mouse anti-phospho-H2AX (Ser139) antibody (1:500; ab26350, Abcam) and rabbit anti-35BP1 (1:500; ab36823, Abcam). Slides were incubated with anti-mouse AlexaFluro 488-conjugated (1:400; ab150113, Abcam) and anti-rabbit AlexaFluro 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 Y-H2AX and 35BP1 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 v.3.1.8. 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**

**Code availability**

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

**References**

1. Kuchenbaecker, K. B. et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA 317, 2402–2416 (2017).

2. 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).

3. 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).

4. 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).

**Received: 15 November 2021; Accepted: 12 September 2022; Published online: 06 October 2022**
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. 24, 1656–1668 (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. Zerbini, A., Macdonald, J. R., Mertico, 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 variants 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, 452–438 (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 germine 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. Moyer, 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. Pharmacogenet. Genomics 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 sulphotransferase SULT1A1 (ST1A3) and arylamine N-acetyltransferase NAT2, with urothelial cancer in a Japanese population. Int. J. Cancer 102, 418–426 (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 influence bladder cancer risk. Chem. Res. Toxicol. 16, 444–450 (2007).
27. 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).
28. Chen, S. & Parmigiani, G. Meta-analysis of BRCA1 and BRCA2 penetrance. J. Clin. Oncol. 25, 1329–1333 (2007).
29. 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. 43, e31–e37 (2005).
30. Loveck, P. K. et al. Identification of BRCA1 missense substitutions that confer partial functional activity: potential moderate risk variants? Breast Cancer Res 9, R82 (2007).
31. Spurdle, A. B. et al. BRCA1 R169Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. J. Med. Genet. 49, 525–532 (2012).
32. 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).
33. 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).
34. 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, 1774–1792 (2015).
35. R Core Team. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2013).

9
Acknowledgements

All the families and clinicians who contribute to the studies; Catherine M. Phelan for her contribution to CIMP 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 Maskelli, 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; Vilus Ruditis and Laïmonas Grinkiūnas; 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 Ejlertsen and Anne-Marie Gerdes for the recruitment and genetic counselling of participants; Alicia Barroso, Rosario Alonso and Guillermo Pitta; 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, Olivia Fuentes, Giuseppe Giannini and the personnel of the Cogentech Tumori Ereditari Alla Mammella), in particular: Bernard Peissel, Dario Zimbalatti, Daniela Zaffaroni, Laura Ottini, Olivia Fuentes, Giuseppe Giannini and the personnel of the Cogentech.

This study was supported by a grant from the New York City Cancer Project and the many clinicians and patients for their contributions to this resource, and the many families who contribute to KONFab; the KOBRA Study Group; Cuila Szabo (National Institute of Public Health, Hungary); and Richard D. Siegel, MD, for his help in initiating and coordinating the study. We thank all families and clinicians who contribute to the studies; Catherine M. Phelan for her contribution to CIMP 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 Maskelli, 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; Vilus Ruditis and Laïmonas Grinkiūnas; 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 Ejlertsen and Anne-Marie Gerdes for the recruitment and genetic counselling of participants; Alicia Barroso, Rosario Alonso and Guillermo Pitta; 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, Olivia Fuentes, Giuseppe Giannini and the personnel of the Cogentech.
Zealand Health Research Council (#17111) and the Mackenzie Charitable Foundation. CIMBA: The CIMBA data management and data analysis were supported by Cancer Research UK grants (A20861 and A20861/003), the National Cancer Institute of the National Institutes of Health, the Netherlands Cancer Institute (KKNL; 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 HealthR 03CA30065, and North California Cancer Center. The study was supported by the Spanish Instituto de Investigaciones Hospitales Clínicos Ramón y Cajal (CIHHF) and the Isabel Zendal. FBBCC: Research Grant Program 20-172-2017. Cancer Research Foundation (project 20-1722017). BCCFC: Research Grant Program 20-172-2017. Cancer Research Foundation (project 20-1722017). BCCFC: Research Grant Program 20-172-2017. Cancer Research Foundation (project 20-1722017). BCCFC: Research Grant Program 20-172-2017. Cancer Research Foundation (project 20-1722017). BCCFC: Research Grant Program 20-172-2017. Cancer Research Foundation (project 20-1722017). BCCFC: Research Grant Program 20-172-2017. Cancer Research Foundation (project 20-1722017).
Christopher Hakkaart, John F. Pearson, Louise Marquart, Joe Dennis, George A. R. Wiggins, Daniel R. Barnes, Bridget A. Robinson, Peter D. Mace, Kristiina Aittomäki, Irene L. Andrilis, Banu K. Arun, Jacopo Azzollini, Judith Balmaña, Rosa B. Barkardottir, Sami Belhadj, Lieke Berger, Julika Borde, Angela R. Bradbury, Joan Brunet, Kristiina Aittomäki, Irene L. Andrulis, Judith Balmaña, Rosa B. Barkardottir, Sami Belhadj, Lieke Berger, Julika Borde, Angela R. Bradbury, Joan Brunet, Kristiina Aittomäki, Irene L. Andrulis, Judith Balmaña, Rosa B. Barkardottir, Sami Belhadj, Lieke Berger, Julika Borde, Angela R. Bradbury, Joan Brunet.
Michael T. Parsons, Inge Sokilde Pedersen, Ana Peixoto, Pedro Perez-Segura, Paolo Peterlongo, Timea Pozca, Paolo Radice, Juliane Ramser, Johanna Rantalainen, Gustavo C. Rodriguez, Karina Ranlund, Efraim H. Rosenberg, Maria Rossing, Rita K. Schmutzler, Payal D. Shah, Saba Sharifi, Priyanka Sharma, Lucy E. Side, Jacques Simard, Christian F. Singer, Katie Snape, Doris Steinemann, Dominique Stoppa-Lyonnet, Christian Sutter, Yen Yen Tan, Manuel R. Teixeira, Soo Hwang Teo, Mads Thomassen, Darcy L. Thull, Marc Tischkowitz, Amanda E. Toland, Alison H. Trainer, Nadine Tung, Klaartje van Engelen, Elizabeth J. van Rensburg, Ana Vega, Alessandra Viel, Lisa Walker, Jeffrey N. Weitzel, Marike R. Wevers, Georgia Chenevix-Trench, Amanda B. Spurdle, Antonis C. Antoniou, & Logan C. Walker.
**GEMO Study Collaborators**
Marie-Agnès Collonge-Rame, Capucine Delnatte, Paul Gesta, Christine Lasset, Clémentine Legrand, Fabienne Lesueur, Véronique Mari, Noura Mebirouk, Emmanuelle Mouret-Fourme, Sophie Nambot & Dominique Stoppa-Lyonnet

**EMBRACE Collaborators**
Jackie Cook, Rosemarie Davidson, Alan Donaldson, Jacqueline Eason, D. Gareth Evans, Debra Frost, Louise Izatt, Zoe Kemp, Fiona Laloo, Hannah Musgrave, Saba Sharif, Lucy E. Side, Katie Snape, Vishakha Tripathi & Lisa Walker

**SWE-BRCA Investigators**
Hans Ehrencrona & Johanna Rantala

**kConFab Investigators**
Georgia Chenevix-Trench & Amanda B. Spurdle

**HEBON Investigators**
Lieke Berger, Marinus J. Blok, Marco Koudijs, John W. M. Martens, Efraim H. Rosenberg, Klaartje van Engelen & Marije R. Wevers