cancers

MDPI

Article

The Contribution of Germline Pathogenic Variants in Breast Cancer Genes to Contralateral Breast Cancer Risk in *BRCA1/BRCA2/PALB2*-Negative Women

Alexey Larionov [1,2,*], Eleanor Fewings [1], James Redman [1], Mae Goldgraben [1], Graeme Clark [1], John Boice [3], Patrick Concannon [4], Jonine Bernstein [5], David V. Conti [6], The WECARE Study Collaborative Group [†] and Marc Tischkowitz [1,*]

[1] Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge CB2 0QQ, UK
[2] School of Water, Energy and Environment, Cranfield University, Bedford MK43 0AL, UK
[3] Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[4] Genetics Institute and Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL 32610, USA
[5] Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
[6] Division of Biostatistics, Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA 90032, USA
[*] Correspondence: alexey.larionov@cranfield.ac.uk (A.L.); mdt33@medschl.cam.ac.uk (M.T.)
[†] Membership of the WECARE Study Collaborative Group is provided in the Acknowledgments. woodsm@mskcc.org.

Simple Summary: As the treatment for breast cancer continues to improve and more women survive their initial diagnosis, there is an increasingly large number of women who are at risk of a second new breast cancer in their lifetimes. However, the hereditary causes of these second breast cancers are not well understood. In this study, we used the latest genetic sequencing technologies to investigate hereditary causes for the second breast cancer in individuals who are known not to have alterations in one of the three main breast cancer genes (*BRCA1*, *BRCA2* and *PALB2*). We analyzed the genetic profiles of selected participants from the WECARE study, one of the largest studies looking at second breast cancers in women. By comparing the genetic profiles of women who have had one breast cancer to similarly matched women who went on to have a second breast cancer, we found that younger women (under 50) with second breast cancers had a higher number of inherited gene alterations compared with those women with one breast cancer. We did not see the same effect in the older women. The results from this study improve our understanding of the hereditary contribution to second breast cancers.

Abstract: Background: Contralateral breast cancer (CBC) is associated with younger age at first diagnosis, family history and pathogenic germline variants (PGVs) in genes such as *BRCA1*, *BRCA2* and *PALB2*. However, data regarding genetic factors predisposing to CBC among younger women who are *BRCA1/2/PALB2*-negative remain limited. Methods: In this nested case-control study, participants negative for *BRCA1/2/PALB2* PGVs were selected from the WECARE Study. The burden of PGVs in established breast cancer risk genes was compared in 357 cases with CBC and 366 matched controls with unilateral breast cancer (UBC). The samples were sequenced in two phases. Whole exome sequencing was used in Group 1, 162 CBC and 172 UBC (mean age at diagnosis: 42 years). A targeted panel of genes was used in Group 2, 195 CBC and 194 UBC (mean age at diagnosis: 50 years). Comparisons of PGVs burdens between CBC and UBC were made in these groups, and additional stratified sub-analysis was performed within each group according to the age at diagnosis and the time from first breast cancer (BC). Results: The PGVs burden in Group 1 was significantly higher in CBC than in UBC (*p* = 0.002, OR = 2.5, 95CI: 1.2–5.6), driven mainly by variants in *CHEK2* and *ATM*. The proportions of PGVs carriers in CBC and UBC in this group were 14.8% and

5.8%, respectively. There was no significant difference in PGVs burden between CBC and UBC in Group 2 (*p* = 0.4, OR = 1.4, 95CI: 0.7–2.8), with proportions of carriers being 8.7% and 8.2%, respectively. There was a significant association of PGVs in CBC with younger age. Metanalysis combining both groups confirmed the significant association between the burden of PGVs and the risk of CBC (*p* = 0.006) with the significance driven by the younger cases (Group 1). Conclusion: In younger *BRCA1/BRCA2/PALB2*-negative women, the aggregated burden of PGVs in breast cancer risk genes was associated with the increased risk of CBC and was inversely proportional to the age at onset.

1. Introduction

Advances in the treatment of breast cancer (BC) over recent decades have substantially improved the survival rates of women with BC, thereby increasing the population of women at risk of a second primary BC. The risk of developing asynchronous contralateral BC (CBC) is higher than the risk of a first BC in the population [1–3], equating to at least 4% at ten years after the first BC [3,4]. The development of CBC is associated with a worse prognosis and other morbidities [5,6]. However, there is no clinical consensus for managing the individual risk of developing CBC, with rates of prophylactic bilateral mastectomies varying significantly across U.S. clinics [7–9]. A number of informatic tools estimating individual CBC risk have been deployed to address this challenge [10–13], but given the lack of data available, their utility and generalizability for accurate clinical decision-making have been limited [13,14].

Risk factors for CBC include young age at the first BC and family history [1,15], indicating that hereditary factors play a role [16]. Treatment for the first BC is also known to modify subsequent CBC risk and has been studied in combination with specific genetic risk factors; examples include the use of tamoxifen in carriers of pathogenic germline variants (PGVs) in *BRCA1* and *BRCA2* [17–19] and the effects of radiation exposure in combination with pathogenic variants in *ATM* [20]. Overall, the use of chemotherapy and hormone therapy has been linked to a decrease in CBC risk by 42% [6,21,22].

Pathogenic germline variants (PGVs) in *BRCA1* and *BRCA2* have been associated with an increased risk of CBC [23–25]. Carriers of PGVs in *BRCA1* and *BRCA2* are highly susceptible to CBC development with a risk of 2.5% per year after diagnosis of the first BC [24], which is at least five times higher than in non-carriers [2,3]. A number of studies have attempted to identify further PGVs associated with CBC risks, including previous WECARE (for Women's Environment, Cancer and Radiation Epidemiology) Study publications that analyzed *CHEK2**1100delC and variants in *ATM* and *PALB2* [20,26–28]. However, the role of PGVs beyond these genes is not yet known.

The WECARE Study is a population-based case-control study that collected phenotype data and blood DNA samples from women with CBC (cases) and individually matched controls with unilateral BC (UBC) [29]. The rich phenotypic data available from the WECARE Study, including demographic, epidemiologic and clinical information, provide the opportunity to study predisposition to CBC, while adjusting for known general risk factors of CBC. A number of previous publications based on the WECARE Study participants examined the joint roles of environmental risk factors, common variants and rare variants in individual genes, primarily focusing on *BRCA1* and *BRCA2* [18,25,27,30–34]. The present study investigated the aggregated burden of PGVs in established BC genes beyond *BRCA1/2* and *PALB2*, focusing on *BRCA1/2/PALB2*-negative participants in the WECARE Study.

2. Materials and Methods

2.1. Study Population, Library Preparation and Sequencing of Samples

The participants with CBC ("cases") and UBC ("controls") were recruited to WE-CARE Study as described earlier [29]. Briefly, all cases and controls were diagnosed before age 55 years from 1985 to 2000 with a first primary invasive breast cancer that had not spread beyond regional lymph nodes. Cases were diagnosed with a second primary invasive or in situ CBC at least one year after a first primary diagnosis in 1986 to 2001. Controls were individually matched to each case on birth year, year of first primary diagnosis, cancer registry and race/ethnicity and were required to have an intact contralateral breast. Controls were assigned a reference date reflecting a cancer-free at-risk period following the first BC equivalent to the interval between first and second diagnoses of the matched cases (latency). Cases and controls were required to have had no other prior or intervening cancers, to have resided in a cancer registry catchment area at both diagnoses/reference date, and to have provided a blood sample. Informed consent was obtained from participants. For this study, one control was matched to each case and known carriers of PGVs in *BRCA1/2/PALB2* were excluded.

For the current study, participants selected were of European ancestry from the U.S., identified through population-based cancer registries covering Iowa, California (Los Angeles and Orange Counties) and Washington State associated with the Surveillance, Epidemiology and End Results (SEER) registry system. This study was conducted in two phases using different sequencing techniques (termed "Group 1" and "Group 2"): whole exome sequencing (WES) was conducted in Group 1 while targeted panel sequencing was used for Group 2 using a custom Ampliseq panel that included 106 genes related to BC and DNA repair (Supplementary Table S1). Across both groups, participants were over-selected for family history of breast cancer in a first-degree relative, and further matched on the dates of diagnosis and follow-up (latency), race and reporting region. Participants with an early age of onset were prioritized during the 1st group selection. Cases and controls were initially individually matched 1:1; however, as some samples failed sequencing, and some of the individual case-control matching was broken, matching factors were adjusted for the analyses along with cytotoxic chemotherapy, hormonal therapy, breast irradiation, the number of pregnancies and the top PCs, as described below. The models were not adjusted for family history of BC because this could be a proxy for carrying pathogenic germline variants. The age distribution between the two groups was different because the earlier onset cases were selected for Group 1 (Group 1: mean age 42; Group 2: mean age 50).

For both groups, DNA was extracted from blood using QIAGEN columns (QIAGEN LLC, USA, Germantown). For Group 1, the WES libraries (125xPE, 24x multiplexed) were prepared using Illumina Nextera Rapid Exome kits (Illumina, Great Abington, Cambridge, UK; FC−140–9001) following the manufacturer's recommendations. Sequencing of WES libraries was performed using Illumina HiSeq−2500 machines (Illumina, Great Abington, Cambridge, UK) and SBS v3 or v4 kits (in CRUK Cambridge Institute Genetic core). Each library was sequenced on 4–6 lanes to reach required depth on targets (68 ± 20, mean $\pm$ SD). For Group 2, the Ampliseq libraries were prepared and sequenced in SMCL genomics core (Department of Medical Genetics, Cambridge University). The custom library was designed using Thermo Fisher Ampliseq Designer tool (Thermo Fisher Scientific, Waltham, USA). Libraries (150PE, 2 pools) were prepared using a protocol adapted from Konig et al., 2015 [35]. Ampliseq samples were sequenced on 2 lanes of HiSeq 4000 Illumina machines (Illumina, Great Abington, Cambridge, UK) to a mean depth on targets of 442 ± 196 (mean $\pm$ SD).

A deidentified study dataset used for the analysis in this publication, which complies with regulatory data sharing restrictions, will be made available upon qualified request to WECARE Study Collaborative Data Repository,

2.2. Alignment, Variant Calling and Annotation

The study flowchart is presented in Supplementary Figure S1. An ethnically close subset of samples (non-Finnish female Europeans, NFFE) from One Thousand Genomes project [36] was used for the joint variant calling with WECARE samples. In addition, these NFFE samples were used for a supplementary supportive analysis to verify that burden of observed PGVs in UBC or CBC was higher than in an unselected population. FASTQ files were demultiplexed and passed through standard QC checks (including FastQC v.0.11.3 and multi-genome alignment [37]). Adaptors and low-quality bases were trimmed using Cutadapt [38]. Alignment and variant calling were performed following GATK Best Practices recommended by the Broad institute at the time of analysis. Reads were aligned to GRCh37 reference genome using BWA MEM (v. 0.7.12) [39]. BAM files from multiple lanes were merged and sorted using samtools (v.1.2) [40]. PCR duplicates were removed in WES, but not in Ampliseq data; diverse QC metrics for alignment and enrichment were then calculated by Picard (https://github.com/broadinstitute/picard accessed on 30 November 2022). FASTQ files for 196 NFFE samples of good quality were selected from One Thousand Genomes project (1KGP) dataset [36]. NFFE FASTQ files were processed by the same pipeline as the WECARE WES data.

Variant calling was performed using GATK (v. 3.4–46 and 3.6–0) [41], and g.vcf files were generated using Haplotype Caller after base quality recalibration and local realignment around indels. Individual g.vcf files were combined into batches of ~100 samples; then, the combined g.vcf-s were used to perform the joint genotyping. Variant calling was performed within the targeted areas only (with 10 bp padding). Variants were filtered by a combination of GATK VQSR and bespoke hard filters, and multiallelic variants were then split to separate lines. GATK down-sampling was suppressed for Ampliseq variant calling and filtering. Variant annotations were added to VCF file using VEP (v.101) [42] and ClinVar (v.20200905) [43]. Along with the standard predicted consequences, the VaP annotations included SIFT (v.5.2.2) [44], PolyPhen (v.2.2.2) [45] and CADD (v.1.6) [46].

Calculations were performed using CRUK CI and Cambridge University high-performance computing clusters. The source data may be requested from the WECARE consortium. The scripts for key steps of the pipeline are available upon request to the WECARE Study Collaborative Data Repository.

2.3. Selecting Genes and Pathogenic Variants

The list of BC susceptibility genes was compiled from the studies of Easton et al. (2015), Hu et al. (2021) and Dorling et al. (2021) [47–49]. It included *BRCA1/2*, *PALB2*, *ATM*, *CHEK2*, *CDH1*, *TP53*, *PTEN*, *STK11*, *NF1*, *NBN*, *RAD51C*, *BARD1* and *RAD51D.* Carriers of PGVs in *BRCA1/2* and *PALB2* were excluded from this analysis. Variants in the remaining 11 genes were analyzed in the WES dataset. *BARD1* and *RAD51D* were not available in the Ampliseq multigene panel, leaving nine BC susceptibility genes for the analysis in the Ampliseq dataset. As the focus of this study was on potentially actionable rare variants with high effect size, rather than on polygenic risks or gene interactions, we used relatively strict criteria in our definition of PGVs:

1. Variants annotated as Pathogenic/Likely pathogenic by ClinVar or
2. Loss of Function variants as predicted by VEP (stop/start gain/loss, frameshift, essential splice sites) or
3. Missense variants that were simultaneously called Deleterious by PolyPhen, Probably Damaging by SIFT and had CADD Phred score > 25.

Of those variants that fulfilled the above criteria, only rare variants (AF < 0.01) with a call rate >0.85 were kept in the analysis. Variants annotated as Benign/Likely benign in ClinVar were excluded.

2.4. Statistical Analysis

The burden of PGVs aggregated across the preselected genes taken together was compared in CBC and UBC. Visual assessment of PGVs burden was performed using bar-plots

showing aggregated allelic frequencies (AF) and proportions of carriers in the compared groups. The aggregated AF was defined as the total count of ALT alleles (AC) over the mean allele number (AN) across the aggregated variants. The statistical significance of differences between CBC and UBC was evaluated by the Burden test as implemented in the *SKAT* R library (v.2.0.1) [50]. Specifically, *p*-values for dichotomous outcome were calculated using efficient resampling, default (linear weighted) kernel and default settings for imputation and weighting. After excluding variants with call rates <0.85, only a small proportion of genotypes required imputation, where the default SKAT-binary imputation method assigned the most likely values observed in the data (0,1,2) to the missing genotypes. The default weighting implemented in SKAT assigns higher weight to rare variants: numerically the weights are calculated using beta-distribution as dbeta(MAF,1,25). As noted above, covariates added into the model included age at the first CBC, latency, number of pregnancies, use of chemo- and hormonal therapies, and breast irradiation for the first BC. Principal components (PCs) for ancestry were calculated using *bigsnpr* R library (v.1.6.1) [51] based on common variants not in linkage disequilibrium across all genes available in WES/panel data. European ancestry of participants was confirmed by projecting WECARE Study samples to PC space of the major 1KGP populations (Supplementary Figures S2A and S3A). Visual assessment of the Scree plots suggested adding the two top PCs as covariates for both Group 1 and Group 2 regression models to account for potential confounding due to residual population stratification (Supplementary Figures S2B and S3B). Three top PCs were added to the POLR model in the additional joined analysis of WES + NFFE after evaluation of the Scree plot for the joined WES + NFFE dataset (Supplementary Figure S4). PCA outliers (if any) were excluded.

Additional subgroup analyses were performed to compare burdens of PGVs in CBC and UBC depending on age or latency (time to CBC) within Group 1 and Group 2 separately. For the age subgroup analysis, study participants were split by median age in each group; the latency threshold for the sub-analysis was set to 5 years. The burden of PGVs within the subgroups was compared using the Fisher test for crude counts and visualized using bar-plots.

Because NFFE controls were unselected for BC, the PGVs prevalence in BC genes is expected to be lower in NFFE than in any WECARE Study sample group (UBC or CBC). Therefore, an additional supportive data check was performed to assess the PGVs burdens between NFFE, UBC and CBC samples. Only variants that had at least 0.85 call rate in both WECARE and NFFE samples were used for this analysis. The NFFE-UBC-CBC trend was visualized by bar-plots for crude counts. Proportional Odds Logit Regression (POLR), as implemented in the *MASS* R package [52], was used for the trend significance assessment (including top PCs into the model).

Metanalysis of summary statistics (*p*-values for PGVs burdens in Group 1 and Group 2) was performed using the METAL package (v.2011–03−25) [53] employing a weighted Z-score to obtain a summary test statistic and *p*-value (i.e., Stouffer's approach).

3. Results

Of the total of 748 samples used for this analysis, 25 samples failed sequencing (15 samples failed WES and 10 samples failed the panel sequencing); 723 samples were successfully sequenced: 334 (172 UBC and 162 CBC) in Group 1 and 389 (194 UBC and 195 CBC) in Group 2.

Table 1 summarizes characteristics of women included in the statistical analysis. The limited sample size precluded analysis of individual variants. At the individual gene level, no genes reached statistical significance after multiple testing correction. In the analysis of PGVs burden over all predefined BC susceptibility genes taken together, the aggregated burden of PGVs in CBC was significantly higher than in UBC in Group 1 (Figure 1A, SKAT Burden test *p* = 0.002). The proportions of PGVs carriers in CBC and UBC in this group were 14.8% and 5.8%, respectively. This was driven mainly by variants in *CHEK2* and *ATM* (Table 2). The full details of PGVs detected in the younger cohort (Group 1, mean age at

first breast cancer = 42 years) are given in Supplementary Table S2. There was no significant difference in the burden of PGVs between CBC and UBC in Group 2 (mean age at diagnosis of the first BC = 50 years), with proportions of carriers being 8.7% and 8.2%, respectively (Figure 1B, SKAT Burden test *p* = 0.42; details of PGVs detected in Group 2 are given in Supplementary Tables S3 and S4).

Table 1. Participant characteristics.

Parameters		Group 1			Group 2		
		UBC	CBC	*p*	UBC	CBC	*p*
Samples Count		172	162	-	194	195	-
Age at the 1st cancer		42 ± 5.5	42 ± 5.4	0.29	50 ± 4.1	50 ± 3.9	0.19
Time to the 2nd cancer or lack of it		6.2 ± 3.5	6.2 ± 3.5	0.88	4.6 ± 2.8	4.6 ± 2.8	0.87
Stage	1	118	119	0.34	143	144	1.00
	2	54	43		51	51	
Estrogen receptors	Neg	47	34	0.31	47	53	0.68
	Pos	92	89		116	108	
	Unknown	33	39		31	34	
Cytotoxic chemotherapy	No	75	83	0.19	99	113	0.19
	Yes	97	79		95	82	
Hormonal therapy	No	115	113	0.55	95	113	0.08
	Yes	56	49		99	82	
	Unknown	1	0		-	-	
Breast irradiation	No	72	89	0.02	63	130	<0.001
	Yes	100	73		131	65	
Number of pregnancies	None	26	38	0.04	32	36	0.78
	1–2	123	113		86	80	
	3 or more	23	11		76	79	
Family history	No	117	97	0.14	133	132	0.98
	Yes	55	65		54	56	
	Unknown	-	-		7	7	

Table notes: *p*-values show significance of differences between contralateral and unilateral breast cancers (CBC and UBC respectively). For continuous data, table shows mean ± std deviation; the significance is estimated by *t*-test. For counts, the significance was estimated by Fisher test for 2 × 2 tables and by Chi square test for 2 × 3 data. CBC: contralateral breast cancer; UBC: unilateral breast cancer.

Table 2. PGVs burden per gene in Group 1.

Gene	Variant Count	UBC			CBC		
		Aggregated AC	Mean AN	Aggregated AF	Aggregated AC	Mean AN	Aggregated AF
ATM	11	7	344.0	0.020	10	323.5	0.031
CHEK2	5	2	344.0	0.006	7	324.0	0.022
NF1	2	0	344.0	0.000	2	324.0	0.006
RAD51D	2	2	343.0	0.006	2	324.0	0.006
NBN	1	0	344	0.0000	2	324	0.0062
CDH1	1	0	344	0.0000	1	324	0.0031
TP53	1	0	336	0.0000	1	320	0.0031
RAD51C	1	0	344	0.0000	1	324	0.0031
Total	24	11	343.6	0.032	26	323.6	0.080

Table notes: Aggregated ALT allele count (aggregated AC) was calculated as sum of ALT allele counts over all variants detected in a gene. Mean allele number (mean AN) was calculated as an arithmetic mean of total allele counts for each variant detected in a gene. For instance, if a gene had two variants, and one of them had AN 345 while the other had AN 346, then the mean AN would be 345.5. Aggregated ALT allele frequency (aggregated AF) was calculated as the total count of ALT alleles (aggregated AC) over the mean allele number (mean AN) across the aggregated variants. No pathogenic variants were identified in the other genes studied in Group 1.
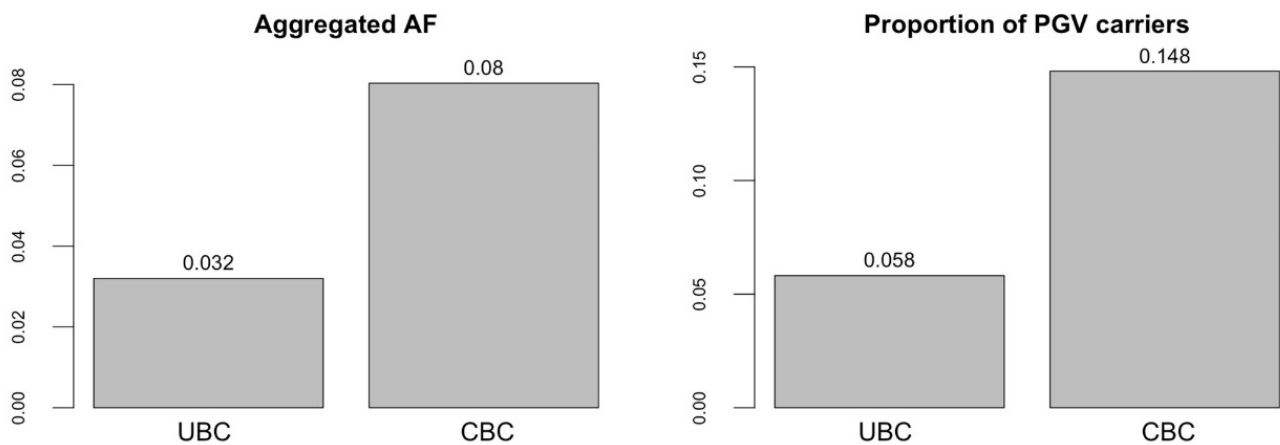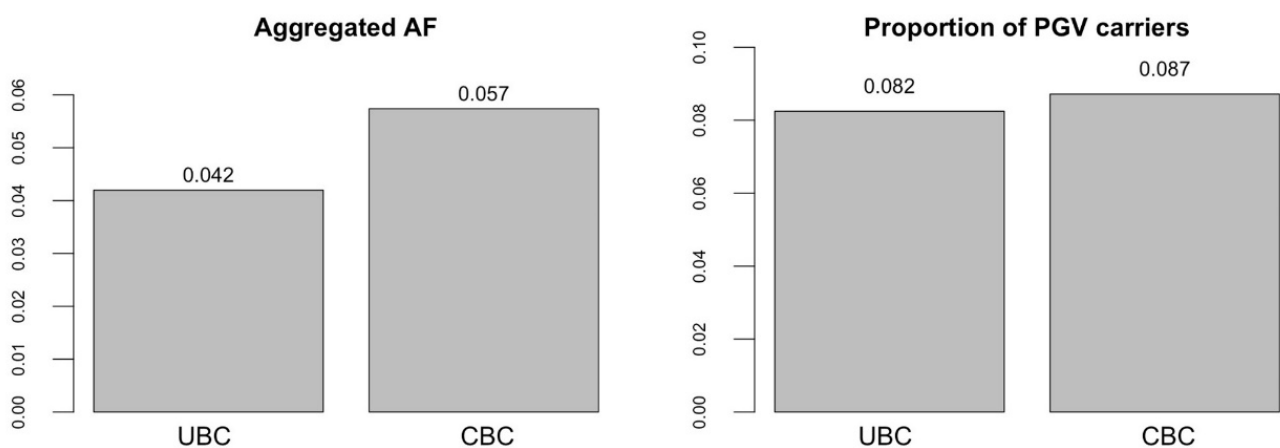
A: Group 1 (mean age 42 ± 5 years)

## B: Group 2 (mean age 50 ± 4 years)

Figure 1. Burden of PGVs in Group 1 and Group 2. Figure notes: In Group 1, the PGVs burden in CBC is significantly higher than in UBC (Burden SKAT test $p = 0.002$). Differences between CBC and UBC in Group 2 are not statistically significant (Burden SKAT test $p = 0.42$). Aggregated ALT allele frequency (aggregated AF) was calculated as the total count of ALT alleles (aggregated AC) over the mean allele number (mean AN) across the aggregated variants. The proportion of PGVs carriers was calculated as the fraction of participants carrying at least one variant in question (homo- or heterozygous). PGV: pathogenic germline variant.

3.1. Subgroup Analyses by Age and Latency

To further investigate the suggested association with age, we performed a stratified subgroup analysis within each of the studied groups separately, splitting them by the median age at the first BC: 43 years in the younger cohort (Group 1) and 51 years in the older cohort (Group 2). The stratified analysis confirmed higher PGVs burdens in CBC compared with UBC in the younger subgroups of participants (Figure 2).

Additionally, we performed a similar stratified subgroup analysis by latency (time between the first and second BCs). No significant difference was observed between participants who developed the second tumor before and after 5 years from the first tumor (Supplementary Figure S5).

A: Group 1 (split by median age: 43 years)
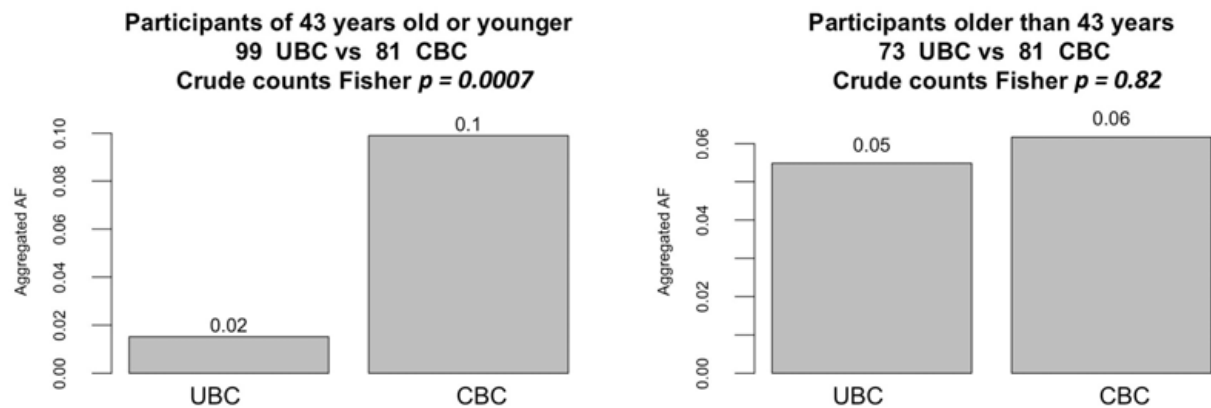
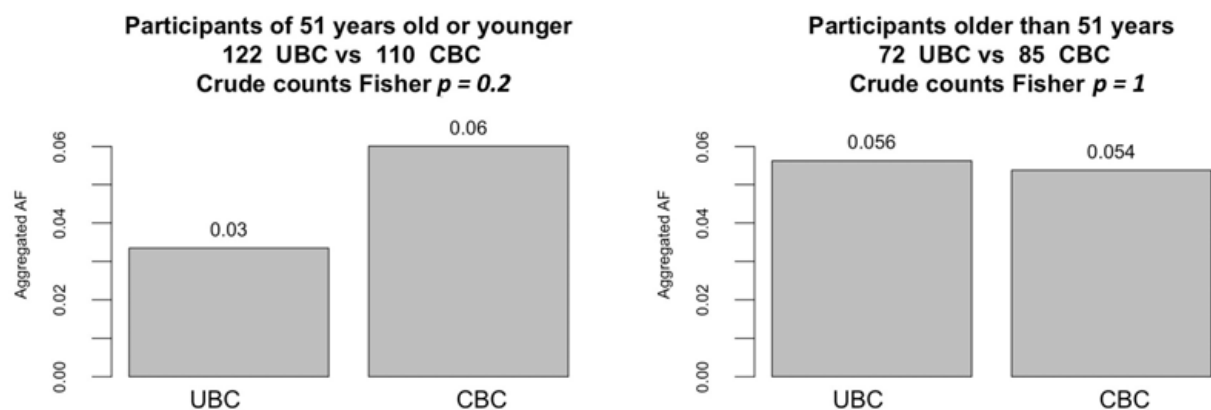B: Group 2 (split by median age: 51 years)

Figure 2. Burden of PGVs stratified by age. Figure note: Each group was split by median age; then, PGVs burdens were evaluated in the subgroups.

3.2. Metanalysis and Comparison of PGVs Burdens with NFFE

As expected, the burden of observed PGVs in NFFE was lower than in either UBC or CBC (Supplementary Figure S6), consistent with the fact that 1KGP NFFE represents a population not selected for BC.

Metanalysis combining data from both Groups 1 and 2 confirmed the higher burden of PGVs in CBC compared with UBC ($p = 0.006$, driven by the participants with younger ages at diagnosis). No significant heterogeneity between the groups was detected in the metanalysis ($p = 0.08$).

4. Discussion

As BC survival rates improve, there is an increasing need to better understand the risk factors for CBC development. This study measured the aggregated burden of PGVs in established BC genes in *BRCA1/2/PALB2*-negative CBC patients who were part of the WECARE Study. We observed that 14.8% of participants with a younger age at CBC diagnosis (Group 1) carried a PGV in at least one of the studied genes, which was significantly higher than in participants with UBC (5.8%) or in the general population. This excess in the aggregated burden of PGVs was mainly driven by variants in *CHEK2* and *ATM* (Table 2). The absence of similar findings in the older group (Group 2) may be explained by the

different biology of BC in women of reproductive age (Group 1: 42 ± 5 years old) and in peri-/postmenopausal age (Group 2: 50 ± 4 years old).

Data on the prevalence of PGVs in CBC beyond *BRCA1/2* and *PALB2* genes are scarce. Yao et al. [54] reported on ~4000 CBC individuals of different ethnicities tested in a single laboratory by panels containing various BC genes. The prevalence of PGVs carriers amongst women with CBC was 3.3% for *CHEK2*, 1.64% for *ATM*, 0.38% for *TP53* and 0.36% for *NBN*. The proportion of *CHEK2* carriers was higher in women with CBC than in women with UBC (*p* < 0.001, driven by individuals of Caucasian origin). However, the study did not have clinical and phenotype data to adjust for treatment, number of pregnancies, and for the fact that some of the first primary BCs, which were used for comparison with CBC, still could develop a second BC later. Fanale et al. (2020) [55] reported results of a panel sequencing for a series of 139 bilateral BCs collected in a single hospital. The study reported 7.9% carriers of PGVs beyond BRCA1/2 and PALB2 genes (five CHEK2, three ATM, two RAD51C and one PTEN variant(s)), which is comparable to the results observed in our study. Of note, the series reported by Fanale et al. [55] included a large proportion of synchronous CBCs (33.1%), and they did not compare the prevalence of PGVs between CBC and UBC or by age at diagnosis.

Our study has several strengths. First, we leveraged a well-characterized study population that allowed the comparison of CBC and UBC, while controlling for multiple known confounding factors. Further, the participants were recruited through population-based cancer registries, rather than in a single hospital series, which enhances generalizability of results. Only early-stage BCs (stages 1 and 2) were included in both UBC and CBC groups, which allowed for longer latency. Including CBC with at least one year between the first and the second cancers ensures that the results predominately apply to the risk of metachronous primary BC. The unique design of the WECARE Study, where UBC controls were individually matched to the CBC cases, as well as the comprehensive phenotype annotations, allowed adjustment for the most important confounding factors influencing the CBC and UBC comparison, such as age at the first BC diagnosis, time to the second BC, treatment, and number of full-term pregnancies.

In spite of these strengths, we were limited by the relatively small sample size: 357 CBC cases and 366 UBC controls altogether, which precluded or limited detailed subgroup analyses or investigation of individual PGVs or genes. This reflects the practical challenges in collecting the dataset where CBC and UBC are matched by age and latency. The limited sample size meant that we had to aggregate variants across multiple genes. Larger BC cohorts will be necessary to assess the individual effects of genes. Two different sequencing techniques were applied to participants with younger and older ages of diagnosis of BC in our study, and while the bioinformatics pipelines were kept as similar as possible, they were not identical.

5. Conclusions

In conclusion, we have estimated the burden of PGVs in established BC risk genes in *BRCA1/BRCA2/PALB2*-negative CBC cases compared with UBC controls, and found an increased aggregated burden of PGVs in established BC-risk genes, which was modified by age. These findings provide further evidence of a strong genetic component of breast cancer etiology, beyond the known high penetrance genes among young women, and demonstrate the importance of screening for PGVs in women diagnosed with BC to identify those who are at increased risk of developing CBC.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/cancers15020415/s1: Supplementary Figure S1: Study flowchart; Figure S2: PCA for Group 1; Figure S3: PCA for Group 2; Figure S4: WECARE-Group1 + NFFE PCA Scree plot; Figure S5: Burden of PGVs stratified by latency; Figure S6: WECARE-Group1 + NFFE PGVs burden; Table S1: PGVs detected in Group 2 samples summarized per gene; Table S2: PGVs detected in Group 1 samples; Table S3: PGVs detected in Group 2 samples; Table S4: PGVs detected in Group 2 samples summarized per gene.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and the study was approved by the local ethics committees and by the WECARE Study Collaborative Group. The WECARE Study biospecimens were collected at multicenter sites between 1999–2004, with Mt Sinai School of Medicine (New York) serving as the coordinating center for the overall study during that time. This manuscript represents scientific work done as an ancillary Whole Exome Study (WECARE WES: Investigating Hereditary Cancer Predisposition, IHCAP) utilizing the 'secondary use of the biospecimens' for a subset of the WECARE Study population. This scientific work obtained institutional regulatory approval at Memorial Sloan Kettering Cancer Center (#17-495 biospecimen research protocol, approved Oct2017), which currently serves as the coordinating center for the overall WECARE Study, and the University of Cambridge (12/EE/0478), where the laboratory work was conducted to generate the WES results.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: A deidentified study dataset used for the analysis in this publication, which complies with regulatory data sharing restrictions, will be made available upon qualified request to the corresponding author (mdt33@medschl.cam.ac.uk).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

BC: breast cancer; CBC: contralateral breast cancer; UBC: unilateral breast cancer; PGVs: pathogenic germline variants; 1KGP: One Thousand Genomes Project; NFFE: non-Finnish female Europeans; SKAT: Sequence Kernel Association Test; POLR: Proportional Odds Logit Regression.

References

1. Bernstein, J.L.; Thompson, W.D.; Risch, N.; Holford, T.R. Risk factors predicting the incidence of second primary breast cancer among women diagnosed with a first primary breast cancer. *Am. J. Epidemiol.* **1992**, *136*, 925–936. [CrossRef] [PubMed]
2. Narod, S.A. Bilateral breast cancers. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 157–166. [CrossRef] [PubMed]
3. Giannakeas, V.; Lim, D.W.; Narod, S.A. The risk of contralateral breast cancer: A SEER-based analysis. *Br. J. Cancer* **2021**, *125*, 601–610. [CrossRef]
4. Reiner, A.S.; John, E.M.; Brooks, J.D.; Lynch, C.F.; Bernstein, L.; Mellemkjaer, L.; Malone, K.E.; Knight, J.A.; Capanu, M.; Teraoka, S.N.; et al. Risk of asynchronous contralateral breast cancer in noncarriers of BRCA1 and BRCA2 mutations with a family history of breast cancer: A report from the Women's Environmental Cancer and Radiation Epidemiology Study. *J. Clin. Oncol.* **2013**, *31*, 433–439. [CrossRef]

5.　Hartman, M.; Czene, K.; Reilly, M.; Adolfsson, J.; Bergh, J.; Adami, H.O.; Dickman, P.W.; Hall, P. Incidence and prognosis of synchronous and metachronous bilateral breast cancer. *J. Clin. Oncol.* **2007**, *25*, 4210–4216. [CrossRef] [PubMed]

6.　Schaapveld, M.; Visser, O.; Louwman, W.J.; Willemse, P.H.; de Vries, E.G.; van der Graaf, W.T.; Otter, R.; Coebergh, J.W.; van Leeuwen, F.E. The impact of adjuvant therapy on contralateral breast cancer risk and the prognostic significance of contralateral breast cancer: A population based study in the Netherlands. *Breast Cancer Res. Treat.* **2008**, *110*, 189–197. [CrossRef]

7.　Nash, R.; Goodman, M.; Lin, C.C.; Freedman, R.A.; Dominici, L.S.; Ward, K.; Jemal, A. State Variation in the Receipt of a Contralateral Prophylactic Mastectomy Among Women Who Received a Diagnosis of Invasive Unilateral Early-Stage Breast Cancer in the United States, 2004–2012. *JAMA Surg.* **2017**, *152*, 648–657. [CrossRef]

8.　Chiesa, F.; Sacchini, V.S. Risk-reducing mastectomy. *Minerva Ginecol.* **2016**, *68*, 544–547.

9.　Liede, A.; Cai, M.; Crouter, T.F.; Niepel, D.; Callaghan, F.; Evans, D.G. Risk-reducing mastectomy rates in the US: A closer examination of the Angelina Jolie effect. *Breast Cancer Res. Treat.* **2018**, *171*, 435–442. [CrossRef]

10.　Basu, N.N.; Ross, G.L.; Evans, D.G.; Barr, L. The Manchester guidelines for contralateral risk-reducing mastectomy. *World J. Surg. Oncol.* **2015**, *13*, 237. [CrossRef]

11.　Chowdhury, M.; Euhus, D.; Onega, T.; Biswas, S.; Choudhary, P.K. A model for individualized risk prediction of contralateral breast cancer. *Breast Cancer Res. Treat.* **2017**, *161*, 153–160. [CrossRef]

12.　O'Donnell, M. Estimating contralateral breast cancer risk. *Curr. Breast Cancer Rep.* **2018**, *10*, 91–97. [CrossRef]

13.　Giardiello, D.; Hooning, M.J.; Hauptmann, M.; Keeman, R.; Heemskerk-Gerritsen, B.A.M.; Becher, H.; Blomqvist, C.; Bojesen, S.E.; Bolla, M.K.; Camp, N.J.; et al. PredictCBC-2.0: A contralateral breast cancer risk prediction model developed and validated in ~200,000 patients. *Breast Cancer Res.* **2022**, *24*, 69. [CrossRef]

14.　Giardiello, D.; Hauptmann, M.; Steyerberg, E.W.; Adank, M.A.; Akdeniz, D.; Blom, J.C.; Blomqvist, C.; Bojesen, S.E.; Bolla, M.K.; Brinkhuis, M.; et al. Prediction of contralateral breast cancer: External validation of risk calculators in 20 international cohorts. *Breast Cancer Res. Treat.* **2020**, *181*, 423–434. [CrossRef]

15.　Reiner, A.S.; Sisti, J.; John, E.M.; Lynch, C.F.; Brooks, J.D.; Mellemkjaer, L.; Boice, J.D.; Knight, J.A.; Concannon, P.; Capanu, M.; et al. Breast Cancer Family History and Contralateral Breast Cancer Risk in Young Women: An Update From the Women's Environmental Cancer and Radiation Epidemiology Study. *J. Clin. Oncol.* **2018**, *36*, 1513–1520. [CrossRef] [PubMed]

16.　Kramer, I.; Hooning, M.J.; Mavaddat, N.; Hauptmann, M.; Keeman, R.; Steyerberg, E.W.; Giardiello, D.; Antoniou, A.C.; Pharoah, P.D.P.; Canisius, S.; et al. Breast Cancer Polygenic Risk Score and Contralateral Breast Cancer Risk. *Am. J. Hum. Genet.* **2020**, *107*, 837–848. [CrossRef]

17.　Gronwald, J.; Tung, N.; Foulkes, W.D.; Offit, K.; Gershoni, R.; Daly, M.; Kim-Sing, C.; Olsson, H.; Ainsworth, P.; Eisen, A.; et al. Tamoxifen and contralateral breast cancer in BRCA1 and BRCA2 carriers: An update. *Int. J. Cancer* **2006**, *118*, 2281–2284. [CrossRef]

18.　Reding, K.W.; Bernstein, J.L.; Langholz, B.M.; Bernstein, L.; Haile, R.W.; Begg, C.B.; Lynch, C.F.; Concannon, P.; Borg, A.; Teraoka, S.N.; et al. Adjuvant systemic therapy for breast cancer in BRCA1/BRCA2 mutation carriers in a population-based study of risk of contralateral breast cancer. *Breast Cancer Res. Treat.* **2010**, *123*, 491–498. [CrossRef] [PubMed]

19.　Brooks, J.D.; Teraoka, S.N.; Malone, K.E.; Haile, R.W.; Bernstein, L.; Lynch, C.F.; Mellemkjaer, L.; Duggan, D.J.; Reiner, A.S.; Concannon, P.; et al. Variants in tamoxifen metabolizing genes: A case-control study of contralateral breast cancer risk in the WECARE study. *Int. J. Mol. Epidemiol. Genet.* **2013**, *4*, 35–48.

20.　Bernstein, J.L.; Haile, R.W.; Stovall, M.; Boice, J.D., Jr.; Shore, R.E.; Langholz, B.; Thomas, D.C.; Bernstein, L.; Lynch, C.F.; Olsen, J.H.; et al. Radiation exposure, the ATM Gene, and contralateral breast cancer in the women's environmental cancer and radiation epidemiology study. *J. Natl. Cancer Inst.* **2010**, *102*, 475–483. [CrossRef]

21.　Bertelsen, L.; Bernstein, L.; Olsen, J.H.; Mellemkjaer, L.; Haile, R.W.; Lynch, C.F.; Malone, K.E.; Anton-Culver, H.; Christensen, J.; Langholz, B.; et al. Effect of systemic adjuvant treatment on risk for contralateral breast cancer in the Women's Environment, Cancer and Radiation Epidemiology Study. *J. Natl. Cancer Inst.* **2008**, *100*, 32–40. [CrossRef]

22.　Langballe, R.; Mellemkjaer, L.; Malone, K.E.; Lynch, C.F.; John, E.M.; Knight, J.A.; Bernstein, L.; Brooks, J.; Andersson, M.; Reiner, A.S.; et al. Systemic therapy for breast cancer and risk of subsequent contralateral breast cancer in the WECARE Study. *Breast Cancer Res.* **2016**, *18*, 65. [CrossRef] [PubMed]

23.　Fletcher, O.; Johnson, N.; Dos Santos Silva, I.; Kilpivaara, O.; Aittomaki, K.; Blomqvist, C.; Nevanlinna, H.; Wasielewski, M.; Meijers-Heijerboer, H.; Broeks, A.; et al. Family history, genetic testing, and clinical risk prediction: Pooled analysis of CHEK2 1100delC in 1,828 bilateral breast cancers and 7030 controls. *Cancer Epidemiol. Biomarkers Prev.* **2009**, *18*, 230–234. [CrossRef] [PubMed]

24.　Metcalfe, K.; Gershman, S.; Lynch, H.T.; Ghadirian, P.; Tung, N.; Kim-Sing, C.; Olopade, O.I.; Domchek, S.; McLennan, J.; Eisen, A.; et al. Predictors of contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. *Br. J. Cancer* **2011**, *104*, 1384–1392. [CrossRef]

25.　Malone, K.E.; Begg, C.B.; Haile, R.W.; Borg, A.; Concannon, P.; Tellhed, L.; Xue, S.; Teraoka, S.; Bernstein, L.; Capanu, M.; et al. Population-based study of the risk of second primary contralateral breast cancer associated with carrying a mutation in BRCA1 or BRCA2. *J. Clin. Oncol.* **2010**, *28*, 2404–2410. [CrossRef] [PubMed]

26.　Tischkowitz, M.; Capanu, M.; Sabbaghian, N.; Li, L.; Liang, X.; Vallee, M.P.; Tavtigian, S.V.; Concannon, P.; Foulkes, W.D.; Bernstein, L.; et al. Rare germline mutations in PALB2 and breast cancer risk: A population-based study. *Hum. Mutat.* **2012**, *33*, 674–680. [CrossRef]

27. Reiner, A.S.; Robson, M.E.; Mellemkjaer, L.; Tischkowitz, M.; John, E.M.; Lynch, C.F.; Brooks, J.D.; Boice, J.D.; Knight, J.A.; Teraoka, S.N.; et al. Radiation Treatment, ATM, BRCA1/2, and CHEK2*1100delC Pathogenic Variants and Risk of Contralateral Breast Cancer. *J. Natl. Cancer Inst.* **2020**, *112*, 1275–1279. [CrossRef]

28. Mellemkjaer, L.; Dahl, C.; Olsen, J.H.; Bertelsen, L.; Guldberg, P.; Christensen, J.; Borresen-Dale, A.L.; Stovall, M.; Langholz, B.; Bernstein, L.; et al. Risk for contralateral breast cancer among carriers of the CHEK2*1100delC mutation in the WECARE Study. *Br. J. Cancer* **2008**, *98*, 728–733. [CrossRef]

29. Bernstein, J.L.; Langholz, B.; Haile, R.W.; Bernstein, L.; Thomas, D.C.; Stovall, M.; Malone, K.E.; Lynch, C.F.; Olsen, J.H.; Anton-Culver, H.; et al. Study design: Evaluating gene-environment interactions in the etiology of breast cancer—The WECARE study. *Breast Cancer Res.* **2004**, *6*, R199–R214. [CrossRef]

30. Borg, A.; Haile, R.W.; Malone, K.E.; Capanu, M.; Diep, A.; Torngren, T.; Teraoka, S.; Begg, C.B.; Thomas, D.C.; Concannon, P.; et al. Characterization of BRCA1 and BRCA2 deleterious mutations and variants of unknown clinical significance in unilateral and bilateral breast cancer: The WECARE study. *Hum. Mutat.* **2010**, *31*, E1200–E1240. [CrossRef]

31. Robson, M.E.; Reiner, A.S.; Brooks, J.D.; Concannon, P.J.; John, E.M.; Mellemkjaer, L.; Bernstein, L.; Malone, K.E.; Knight, J.A.; Lynch, C.F.; et al. Association of Common Genetic Variants With Contralateral Breast Cancer Risk in the WECARE Study. *J. Natl. Cancer Inst.* **2017**, *109*, djx051. [CrossRef] [PubMed]

32. Sisti, J.S.; Bernstein, J.L.; Lynch, C.F.; Reiner, A.S.; Mellemkjaer, L.; Brooks, J.D.; Knight, J.A.; Bernstein, L.; Malone, K.E.; Woods, M.; et al. Reproductive factors, tumor estrogen receptor status and contralateral breast cancer risk: Results from the WECARE study. *Springerplus* **2015**, *4*, 825. [CrossRef] [PubMed]

33. Figueiredo, J.C.; Brooks, J.D.; Conti, D.V.; Poynter, J.N.; Teraoka, S.N.; Malone, K.E.; Bernstein, L.; Lee, W.D.; Duggan, D.J.; Siniard, A.; et al. Risk of contralateral breast cancer associated with common variants in BRCA1 and BRCA2: Potential modifying effect of BRCA1/BRCA2 mutation carrier status. *Breast Cancer Res. Treat.* **2011**, *127*, 819–829. [CrossRef]

34. Poynter, J.N.; Langholz, B.; Largent, J.; Mellemkjaer, L.; Bernstein, L.; Malone, K.E.; Lynch, C.F.; Borg, A.; Concannon, P.; Teraoka, S.N.; et al. Reproductive factors and risk of contralateral breast cancer by BRCA1 and BRCA2 mutation status: Results from the WECARE study. *Cancer Causes Control* **2010**, *21*, 839–846. [CrossRef] [PubMed]

35. Konig, K.; Peifer, M.; Fassunke, J.; Ihle, M.A.; Kunstlinger, H.; Heydt, C.; Stamm, K.; Ueckeroth, F.; Vollbrecht, C.; Bos, M.; et al. Implementation of Amplicon Parallel Sequencing Leads to Improvement of Diagnosis and Therapy of Lung Cancer Patients. *J. Thorac. Oncol.* **2015**, *10*, 1049–1057. [CrossRef]

36. Clarke, L.; Fairley, S.; Zheng-Bradley, X.; Streeter, I.; Perry, E.; Lowy, E.; Tasse, A.M.; Flicek, P. The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res.* **2017**, *45*, D854–D859. [CrossRef]

37. Hadfield, J.; Eldridge, M.D. Multi-genome alignment for quality control and contamination screening of next-generation sequencing data. *Front. Genet.* **2014**, *5*, 31. [CrossRef]

38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]

39. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997v1302.

40. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [CrossRef]

41. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef] [PubMed]

42. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef] [PubMed]

43. Landrum, M.J.; Chitipiralla, S.; Brown, G.R.; Chen, C.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Kaur, K.; Liu, C.; et al. ClinVar: Improvements to accessing data. *Nucleic Acids Res.* **2020**, *48*, D835–D844. [CrossRef]

44. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [CrossRef]

45. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [CrossRef]

46. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **2019**, *47*, D886–D894. [CrossRef]

47. Easton, D.F.; Pharoah, P.D.; Antoniou, A.C.; Tischkowitz, M.; Tavtigian, S.V.; Nathanson, K.L.; Devilee, P.; Meindl, A.; Couch, F.J.; Southey, M.; et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **2015**, *372*, 2243–2257. [CrossRef] [PubMed]

48. Hu, C.; Hart, S.N.; Gnanaolivu, R.; Huang, H.; Lee, K.Y.; Na, J.; Gao, C.; Lilyquist, J.; Yadav, S.; Boddicker, N.J.; et al. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N. Engl. J. Med.* **2021**, *384*, 440–451. [CrossRef]

49. Breast Cancer Association, C.; Dorling, L.; Carvalho, S.; Allen, J.; Gonzalez-Neira, A.; Luccarini, C.; Wahlstrom, C.; Pooley, K.A.; Parsons, M.T.; Fortuno, C.; et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N. Engl. J. Med.* **2021**, *384*, 428–439. [CrossRef]

50. Wu, M.C.; Lee, S.; Cai, T.; Li, Y.; Boehnke, M.; Lin, X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **2011**, *89*, 82–93. [CrossRef] [PubMed]

51.   Prive, F.; Luu, K.; Blum, M.G.B.; McGrath, J.J.; Vilhjalmsson, B.J. Efficient toolkit implementing best practices for principal component analysis of population genetic data. *Bioinformatics* **2020**, *36*, 4449–4457. [CrossRef] [PubMed]
52.   Venables, W.N.; Ripley, B.D. *Modern applied statistics with S-PLUS*, 4th ed.; Springer: Berlin/Heidelberg, Germany, 2002.
53.   Willer, C.J.; Li, Y.; Abecasis, G.R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **2010**, *26*, 2190–2191. [CrossRef] [PubMed]
54.   Yao, K.K.; Clifford, J.; Li, S.; LaDuca, H.; Hulick, P.; Gutierrez, S.; Black, M.H. Prevalence of Germline Pathogenic and Likely Pathogenic Variants in Patients With Second Breast Cancers. *JNCI Cancer Spectr.* **2020**, *4*, pkaa094. [CrossRef]
55.   Fanale, D.; Incorvaia, L.; Filorizzo, C.; Bono, M.; Fiorino, A.; Calo, V.; Brando, C.; Corsini, L.R.; Barraco, N.; Badalamenti, G.; et al. Detection of Germline Mutations in a Cohort of 139 Patients with Bilateral Breast Cancer by Multi-Gene Panel Testing: Impact of Pathogenic Variants in Other Genes beyond BRCA1/2. *Cancers* **2020**, *12*, 2415. [CrossRef] [PubMed]