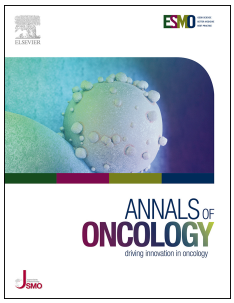
Journal Pre-proof



Analysis of rare disruptive germline mutations in 2,135 enriched *BRCA*-negative breast cancers excludes additional high-impact susceptibility genes

C. Loveday, A. Garrett, P. Law, S. Hanks, E. Poyastro-Pearson, J.W. Adlard, J. Barwell, J. Berg, A.F. Brady, C. Brewer, C. Chapman, J. Cook, R. Davidson, A. Donaldson, F. Douglas, L. Greenhalgh, A. Henderson, L. Izatt, A. Kumar, F. Lalloo, Z. Miedzybrodzka, P.J. Morrison, J. Paterson, M. Porteous, M.T. Rogers, L. Walker, The Breast and Ovarian Cancer Susceptibility Collaboration, D. Eccles, D.G. Evans, K. Snape, H. Hanson, R.S. Houlston, C. Turnbull

PII: S0923-7534(22)04143-6

DOI: https://doi.org/10.1016/j.annonc.2022.09.152

Reference: ANNONC 1099

To appear in: *Annals of Oncology*

Received Date: 3 March 2022

Revised Date: 15 August 2022

Accepted Date: 1 September 2022

Please cite this article as: Loveday C, Garrett A, Law P, Hanks S, Poyastro-Pearson E, Adlard JW, Barwell J, Berg J, Brady AF, Brewer C, Chapman C, Cook J, Davidson R, Donaldson A, Douglas F, Greenhalgh L, Henderson A, Izatt L, Kumar A, Lalloo F, Miedzybrodzka Z, Morrison PJ, Paterson J, Porteous M, Rogers MT, Walker L, The Breast and Ovarian Cancer Susceptibility Collaboration, Eccles D, Evans DG, Snape K, Hanson H, Houlston RS, Turnbull C, Analysis of rare disruptive germline mutations in 2,135 enriched *BRCA*-negative breast cancers excludes additional high-impact susceptibility genes, *Annals of Oncology* (2022), doi: https://doi.org/10.1016/j.annonc.2022.09.152.

Tables

(a)

Gene effect size	Published population frequency of gene (PTVs)		Power to detect (%) in study of 2,135 cases versus 51,377 controls			
			Enriched		Non-enriched	
	Population MAF	Population frequency	DNA repair analysis	WES analysis	DNA repair analysis	WES analysis
BRCA1-like (OR=10.6)	1.01E-03	1 in 494	100.0%	100.0%	99.8%	98.2%
BRCA2-like (OR=5.9)	1.63E-03	1 in 306	100.0%	100.0%	94.0%	80.9%
PALB2-like (OR=5)	7.49E-04	1 in 667	100.0%	100.0%	31.0%	12.2%
CHEK2-like (OR=2.5)	3.07E-03	1 in 163	99.2%	95.6%	22.0%	6.1%
ATM-like (OR=2.1)	1.43E-03	1 in 350	31.1%	11.5%	1.6%	0.2%

(b)

Gene effect size	Population frequency detectable with 90% power			
	DNA repair gene analysis		WES analysis	
	Population MAF	Population frequency	Population MAF	Population frequency
BRCA1-like (OR=10.6)	4.30E-05	1 in 11628	5.59E-05	1 in 8945
BRCA2-like (OR=5.9)	1.70E-04	1 in 2924	2.23E-04	1 in 2242

PALB2-like (OR=5)	2.60E-04	1 in 1946	3.39E-04	1 in 1475
CHEK2-like (OR=2.5)	1.90E-03	1 in 260	2.62E-03	1 in 191
ATM-like (OR=2.1)	3.70E-03	1 in 135	4.98E-03	1 in 100
Gene effect size	Population frequency detectable with 80% power			
	DNA repair gene analysis		WES analysis	
	Population MAF	Population frequency	Population MAF	Population frequency
BRCA1-like (OR=10.6)	3.56E-05	1 in 14045	4.63E-05	1 in 10799
BRCA2-like (OR=5.9)	1.35E-04	1 in 3704	1.82E-04	1 in 2747
PALB2-like (OR=5)	2.04E-04	1 in 2451	2.81E-04	1 in 1779
CHEK2-like (OR=2.5)	1.54E-03	1 in 325	2.16E-03	1 in 231
ATM-like (OR=2.1)	2.98E-03	1 in 168	4.18E-03	1 in 120

Table 1: Power analysis of current study.

(a) Power afforded by study of 2,135 cases and 51,377 controls to discover genes of the OR/MAF profile of *BRCA1*, *BRCA2*, *PALB2*, *CHEK2* and *ATM* with Bonferroni correction undertaking targeted (*N* = 286 DNA repair genes) or exome-wide analysis (*N* = 19,651 genes), comparing use of enriched or unenriched cases (b) Gene population frequency ($MAF_{combined)}$ for genes of equivalent effect size to *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*

and *ATM* for which this study of 2135 enriched cases/51377 controls had 90%/80% power to detect. OR for *BRCA1*, *BRCA2*, *PALB2*, *CHEK2* and *ATM* are derived from Dorling et al.[25]

Haller 0125R2

Title Page

Article Type

Original article

Title

Analysis of rare disruptive germline mutations in 2,135 enriched *BRCA*-negative breast cancers excludes additional high-impact susceptibility genes

Authors List

C. Loveday[1], A. Garrett[1], P. Law[1], S. Hanks[1], E. Poyastro-Pearson[1], J.W. Adlard[2], J. Barwell[3], J. Berg[4], A.F. Brady[5], C. Brewer[6], C. Chapman[7], J. Cook[8], R. Davidson[9], A Donaldson[10], F. Douglas[11], L. Greenhalgh[12], A. Henderson [13], L. Izatt[14], A. Kumar[15], F. Lalloo[16], Z. Miedzybrodzka[17], P.J. Morrison[18], J. Paterson[19], M. Porteous[20], M.T. Rogers[21], L. Walker[22], The Breast and Ovarian Cancer Susceptibility Collaboration[23], D. Eccles[24], D.G. Evans[25], K. Snape[1,26],H. Hanson[1,26], R.S. Houlston[1], C. Turnbull[1,27]

Affiliations List

1. Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK.
2. Yorkshire Regional Genetics Service, St James's University Hospital, Leeds, UK.

3. Leicestershire Genetics Centre, University Hospitals of Leicester National Health Service (NHS) Trust, Leicester, UK.
4. Division of Medical Sciences, Human Genetics, University of Dundee, Dundee, UK.
5. North West Thames Regional Genetics Service, Kennedy Galton Centre, London, UK.
6. Peninsula Regional Genetics Service, Royal Devon & Exeter Hospital, Exeter, UK.
7. West Midlands Regional Genetics Service, Birmingham Women's Hospital, Birmingham, UK.
8. Sheffield Regional Genetics Service, Sheffield Children's NHS Foundation Trust, Sheffield, UK.
9. West of Scotland Regional Genetics Service, Ferguson Smith Centre for Clinical Genetics, Glasgow, UK.
10. South Western Regional Genetics Service, University Hospitals of Bristol NHS Foundation Trust, Bristol, UK.
11. Northern Genetics Service, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.
12. Cheshire and Merseyside Clinical Genetics Service, Alder Hey Children's NHS Foundation Trust, Liverpool, UK.
13. Northern Genetics Service (Cumbria), Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK.
14. South East Thames Regional Genetics Service, Guy’s and St. Thomas NHS Foundation Trust, London, UK.
15. North East Thames Regional Genetics Service, Great Ormond St. Hospital, London, UK.
16. University Department of Medical Genetics & Regional Genetics Service, St. Mary's Hospital, Manchester, UK.

17. University of Aberdeen and North of Scotland Clinical Genetics Service, Aberdeen Royal Infirmary, Aberdeen, UK.
18. Belfast Health and Social Care (HSC) Trust & Department of Medical Genetics, Northern Ireland Regional Genetics Service, Queen's University Belfast, Belfast, UK.
19. East Anglian Regional Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.
20. South East of Scotland Clinical Genetics Service, Western General Hospital, Edinburgh, UK.
21. All Wales Medical Genetics Service, University Hospital of Wales, Cardiff, UK.
22. Oxford Regional Genetics Service, Oxford Radcliffe Hospitals NHS Trust, Oxford, UK.
23. Individual collaborators and their affiliations are listed in the Appendix.
24. Faculty of Medicine, University of Southampton, Southampton University Hospitals NHS Trust, Southampton, UK.
25. University Department of Medical Genetics & Regional Genetics Service, St. Mary's Hospital, Manchester, UK.
26. South West Thames Regional Genetics Service, St. George's Hospital, London, UK.
27. Royal Marsden NHS Foundation Hospital, London, UK.

Manuscript social media information

@clare__turnbull @LovedayChey @DrAliceGarrett @ICR_London www.icr.ac.uk

Corresponding Author

Prof. Clare Turnbull, Division of Genetics and Epidemiology, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, SM2 5NG, UK, 00 44 208 722 4485, clare.turnbull@icr.ac.uk

Manuscript Word Count

3,566

Abstract

Background

Breast cancer has a significant heritable basis, of which approximately 60% remains unexplained. Testing for *BRCA1*/*BRCA2* offers useful discrimination of breast cancer risk within families, and identification of additional breast cancer susceptibility genes could offer clinical utility.

Patients and methods

We included 2,135 invasive breast cancer cases recruited via the BOCS study, a retrospective UK study of familial breast cancer. Eligibility criteria: female, *BRCA*-negative, white European ethnicity, and one of: i) breast cancer family history, ii) bilateral disease, iii) young age of onset (<30 years), iv) concomitant ovarian cancer.

We undertook exome sequencing of cases and performed gene-level burden testing of rare damaging variants against those from 51,377 ethnicity-matched population controls from gnomAD.

Results

159/2135 (7.4%) cases had a qualifying variant in an established breast cancer susceptibility gene, with minimal evidence of signal in other cancer susceptibility genes. Known breast cancer susceptibility genes *PALB2*, *CHEK2* and *ATM* were the only genes to retain statistical significance after correcting for multiple testing. Due to the enrichment of hereditary cases in the series, we had good power (>80%) to detect a gene of *BRCA1*-like risk (odds ratio = 10.6) down to a population minor allele frequency of 4.6×10^{-5} (1 in 10,799, less than one tenth that of *BRCA1*)and of *PALB2*-like risk (odds ratio = 5.0) down to a population minor allele frequency of 2.8×10^{-4} (1 in 1,779, less than half that of *PALB2*). Power was lower for identification of novel moderate penetrance genes (odds ratio = 2-3) like *CHEK2* and *ATM*.

Conclusions

This is the largest case-control whole-exome analysis of enriched breast cancer published to date. Whilst additional breast cancer susceptibility genes likely exist, those of high penetrance are likely to be of very low mutational frequency. Contention exists regarding the clinical utility of such genes.

Keywords

Breast cancer

Genetic susceptibility

Cancer susceptibility genes

Whole exome sequencing

Rare variant burden testing

Highlights

We report the largest exome sequencing study of hereditary breast cancer to date, comparing 2,135 cases to 51,377 controls.

We demonstrate that a novel breast cancer gene (odds ratio ≥ 5) is unlikely to exist at any appreciable population frequency.

We had good power (>80%) to detect any *PALB2*-like breast cancer susceptibility genes (odds ratio = 5) existing at population frequency 1/1,779 (< half that of *PALB2*).

Our insights into the architecture of breast cancer susceptibility give context to the negative findings of the last decade.

Introduction

Female breast cancer is now the most common cancer with 2.3 million cases diagnosed annually worldwide and a lifetime risk of ~15%[1,2]. The heritable basis of the disease is well established and evidenced by the increased risk in relatives of cases[3,4]. It is now well recognised that the genetic architecture of breast cancer susceptibility encompasses a broad spectrum, from common polymorphisms individually conferring small risks through to breast cancer susceptibility genes (BCSGs; Supplementary Table S1) characterized by multiple disparate mutations. BCSGs are typically categorised by their disease penetrance, for example as very high penetrance (odds ratio [OR] ≥ 10), high (OR ≥ 5) or moderate/intermediate (OR ≥ 2)[5].

It is estimated that approximately 41% of the heritable risk of breast cancer has been deciphered[6]. Of the 23% attributable to rare variants, the majority is accounted for by *BRCA1* and *BRCA2* (17%), with other established BCSGs (*PALB2, ATM, CHEK2, TP53, STK11, PTEN, NF1, CDH1, BARD1, RAD51C, RAD51D*) collectively accounting for about 6%[7]. It is estimated that 18% of genetic risk of breast cancer resides in common variants identified by genome-wide association studies (GWAS), with statistical modelling suggesting a further 23% is potentially tractable by larger GWAS[8]. Of the 59% of unexplained heritable risk, over half (36%) is thus potentially ascribable to hitherto unidentified rare variants. Whilst many of these variants may be non-coding and/or of low penetrance, there is considerable interest in the component that may be enshrined in further susceptibility genes of moderate to high penetrance which are both mechanistically important as well as potentially relevant to the clinical management of patients and their families. Numerous studies have used next generation sequencing (NGS)

technologies to search this space for novel BCSGs (Supplementary Table S2), though few if any convincing candidates to date have been identified.

To look for novel BCSGs, we performed germline whole exome sequencing (WES) of 2,135 *BRCA1/BRCA2*-negative genetically enriched breast cancer cases. We used a 3-phase analytical strategy (Figure 1). Firstly, we determined the contribution of pathogenic variants in established autosomal dominant cancer susceptibility genes (CSGs; *N* = 86 genes). Secondly, we assessed the contribution of candidate genes implicated on the basis of involvement in DNA repair (*N* = 276 genes), and/or a known oncogenomic role in cancer (*N* = 686 genes) by performing burden analyses collapsing at gene-level. Finally, we performed exome-wide rare variant burden testing, undertaking correction for cumulative performance of multiple tests.

Methods

Subjects and data

We included *BRCA1*/*BRCA2* negative women with a diagnosis of pathologically confirmed invasive breast cancer (*N* = 2,430), recruited to the ICR Breast and Ovarian Cancer Susceptibility (BOCS) study. Eligibility criteria included one of: (i) a significant family history of breast cancer (breast cancer in at least one first degree-relative or two second degree relatives); ii) bilateral breast cancer; iii) early-onset breast cancer (< 30 years); iv) concomitant ovarian cancer[9]. Cases were assigned a 'genetic enrichment score' based on allele sharing to reflect the degree of personal and family history of breast cancer as follows: bilateral disease in the proband received a score of 1, with additional points assigned for disease in first degree relatives (0.5 each), second-degree relatives (0.25 each) and third-degree relatives (0.125 each). We used a score of ≥ 1.5 to define a set of 'high-risk' cases (*N* = 855).

The proportion of cases of self-reported Ashkenazi Jewish ancestry was < 1%.

Written informed consent was obtained from all participants and the research was approved by the London Multicentre Research Ethics Committee (MREC/01/2/18).

For controls, we used WES data from 51,377 non-Finnish European non-cancer individuals from the Genome Aggregation Database (gnomAD) (date of accession 21/05/2020).

Whole exome sequencing and quality control

Germline DNA from cases underwent WES using Illumina technology. Data processing, variant calling, and annotation were performed according to Genome Analysis Toolkit (GATK) best practices. We excluded cases based on quality, cryptic familial duplicates, presence of occult *BRCA1*/*BRCA2* mutations and ancestry analysis using principal components (Supplementary Figure S1, Supplementary Table S3).

Case/control variant set alignment

Variant level exclusions are described in Supplementary Table S4 and Supplementary Methods. Analyses were restricted to rare variants (MAF < 0.5%). We controlled against spurious inflation of the burden test statistic by filtering the bottom 5th centile of variants based on the GATK variant quality metric QualByDepth (QD) and restricting analyses to base positions with comparable levels of coverage (Supplementary Figure S2, Supplementary Figure S3, Supplementary Table S5, Supplementary Methods)[10].

Statistical analyses

We performed collapsed gene-level burden analyses in: i) autosomal dominantly acting CSGs (*N* = 86)[11, 12]; ii) DNA repair genes (N = 276)[13]; iii) somatic cancer driver genes (*N* = 686 genes); iv) the exome (*N* = 19,651 genes) (Supplementary Table S6).

A single class of test was performed for the CSGs, comprising high confidence protein truncating variants (HC-PTVs, as defined by LOFTEE)[15] plus variants (regardless of variant consequence or LOFTEE annotation) classified as Pathogenic/Likely Pathogenic (P/LP) on ClinVar with one or more stars using The American College of Medical Genetics and Genomics (ACMG) criteria[16]. A single class of test was also performed for missense/in-frame variants at recurrently somatically mutated residues in cancer driver genes (as defined in oncoKB; Supplementary Table S7). For the DNA repair/exome-wide analyses, two classes of test were performed: Class I (HC-PTVs only) and Class II (HC-PTVs + missense with REVEL score > 0.7)[17, 18]. The proportion of REVEL qualifying missense variants that were also pathogenic in ClinVar in known CSGs was 9/172 (5.2%).

Association was measured using a Fishers exact test (2-sided) comparing case/control proportions with a qualifying variant. For declaring statistical significance, we imposed a Bonferroni correction of $P < 0.05$ for the number tests performed, accounting for cumulative testing in candidate-based followed by exome-wide analyses. For replication, we used WES data from UK Biobank comprising 15,017 female breast cancer cases and 199,479 healthy women

with no previous history of cancer, assigning $P < 0.05$ as nominal evidence of replication[19]. Analyses were carried out in R (v4.0.3).

Calculation of study power

Disease allele frequency in controls was taken as the baseline, and the frequency in cases was determined by a weighted average of the predicted enrichment in cases with bilateral disease and/or an affected first degree relative[20]. A Fisher's test (2-sided) was performed for each sampling of cases and controls, performed 10,000 times for each frequency/relative risk combination.

Additional methodological information is available in the Supplementary Methods.

Results

The final case dataset comprised 2,135 *BRCA*-negative breast cancer cases, of which 838 (39%) had bilateral disease, 1628 (76%) had a significant family history of breast cancer (at least one first degree-relative or two second degree relatives), 183 (9%) had breast cancer diagnosed ≤ 30 years, and 261 (12%) had concomitant ovarian cancer (Figure 2). The cases were enriched for bilaterality, family history and early onset disease compared to unselected cases from the literature (Supplementary Figure S4)[21, 22]. Mean age at first diagnosis was 48 years (compared to the UK average of > 60 years[1]).

Frequency of pathogenic germline variants in cancer susceptibility genes (CSGs)

A total of 159/2135 (7.4%) of cases had a qualifying variant (HC-PTV and/or ClinVar P/LP) in one of the thirteen established BCSGs, including 44 (2.1%) in *PALB2*, 41 (1.9 %) in *ATM* and 53 (2.5%) in *CHEK2* (Figure 3, Supplementary Table S8). Eleven cases (0.5%) had a variant in genes associated with estrogen receptor (ER) negative disease: *BARD1* (5, 0.2%), *RAD51C* (4, 0.2%), *RAD51D* (2, 0.1%). Eleven cases (0.5%) had a variant in a rare pleiomorphic cancer susceptibility gene, comprising *TP53* (7, 0.3%), *CDH1* (1, 0.05%), *NF1* (2, 0.1%) and *PTEN* (1, 0.05%). There was no significant difference in the age of breast cancer diagnosis, burden of family history or frequency of bilateral disease between cases with and without a qualifying variant (P > 0.05 in relevant statistical test: see Supplementary Methods).

There were no significant associations with breast cancer amongst the other CSGs, though there was an appreciable case frequency of qualifying variants in some, notably *MSH6* (10

cases) and *BRIP1* (5 cases), genes previously proposed as BCSGs (Figure 3, Supplementary Table S8)[23, 24]. Excluding the thirteen established BCSGs, the frequency of qualifying variants in CSGs was not significantly elevated in cases (53/2135 or 2.5%) compared to controls (1365/51377 or 2.7%) (P = 0.68).

Candidate gene burden testing

Cognisant of the impact of correction for multiple testing, we firstly considered targeted candidacy-based approaches. On the functional basis of genes implicated to date in breast cancer susceptibility, we first considered 276 genes involved in DNA repair and performed burden analysis for rare variants (MAF < 0.5%) of Class I (HC-PTVs only) and Class II (HC-PTVs and deleterious nonsynonymous variants). *PALB2*, *CHEK2* and *ATM* emerged from this analysis as strongly associated (Figure 3, Supplementary Table S8). *WRN* was the highest ranked (non-BCSG) DNA repair gene by P value, driven particularly by the HC-PTV frequency, but the association did not retain significance following Bonferroni correction (Class I test: 10/2135, 0.47% cases vs 73/51377, 0.14% controls, P = 0.002). There was only nominal statistical support on querying of the UK Biobank data for association of this gene with breast cancer, (p=0.0087), which again did not withstand correction for multiple testing (Supplementary Table S9). Excluding the known BCSGs, collapsing across the remaining 263 genes involved in DNA repair the signal of association was not significant on analysis of variants of Class I (316/2135 versus 7130/51377, *P* = 0.23) or Class II (756/2135 versus 18048/51377, *P* = 0.8).

Next, on the basis of exemplar genes associated with both autosomal dominant cancer susceptibility and somatic oncogenic mechanisms (e.g. *RET* mutations predisposing to thyroid cancer and *KIT* mutations predisposing to gastrointestinal stromal tumor [GIST]), we considered 686 cancer driver genes for which recurrent somatic hotspot mutations have been reported. On burden analysis for rare missense/in-frame variants (MAF < 0.5%) overlapping with known oncogenic residues (Supplementary Figure S5, Supplementary Table S8), there were no significant associations after correcting for multiple testing.

Exome-wide rare-variant burden analysis

We performed exome-wide burden testing, again collapsing at gene level for association of rare variants (MAF < 0.5%) of each of Class I and Class II. Again, only for *PALB2* (classes I and II), *CHEK2* (classes I and II) and *ATM* (class II only) were the associations significant following Bonferroni correction (Figure 3, Supplementary Table S8). For all three genes, the odds ratios for breast cancer risk with class I variants were substantially inflated compared to published estimates from unselected breast cancers[25], commensurate with the inflation predicted on the basis of the genetic enrichment in our cases (Supplementary Table S8). Amongst the top twenty associations ranked by smallest P value were *C20orf141* (Class I and Class II) and *EXPH5* (Class I and Class II), but the associations were not significant following Bonferroni correction. Furthermore, neither these genes nor any others in the top 20 associations showed evidence of replication in UK Biobank ($P > 0.1$; Supplementary Table S9).

Notable amongst the top 20 ranked gene associations in this analysis was *PPM1D*. Erroneously reported originally as a BCSG, *PPM1D* is characterized by somatic mosaic mutations arising in hematopoietic cells in response to chemotherapy[26-28]. Consistent with somatic mosaicism in blood cells, the *PPM1D* VAFs detected in this study were skewed downwards for protein truncating variants compared to synonymous variants (Kolmogorov-Smirnov *P* value = 0.11, Mood's median *P* value = 0.17; Supplementary Figure S6).

There was likewise no evidence of association for any genes on subsequent restriction to a smaller set of 855 'high-risk' cases (see Methods, Supplementary Table S8).

On gene-set enrichment analysis of gene ontology (GO) terms and KEGG pathways amongst genes with a Class I/Class II test P value ≤ 0.01 [29, 30], multiple gene groups showed significant associations, primarily related to DNA replication and repair, but once corrected for the signal of association for constituent known BCSGs no gene groups showed significant residual association (Supplementary Table S10).

Power analyses

Recent large-scale genetic epidemiologic analyses have confirmed *BRCA1* as being of almost twice the penetrance but much lower mutational frequency than *BRCA2* (*BRCA1*: OR = 10.6, population $MAF_{combined}$ = 0.10% or 1 in 494; *BRCA2*: OR = 5.9, population $MAF_{combined}$ = 0.16% or 1 in 306)[25]. However, *PALB2* (OR = 5.0, population $MAF_{combined}$ = 0.07% or 1 in 667) has substantially

lower 'impact profile' (penetrance x mutational frequency) with risk in the rough range of *BRCA2* but a mutational frequency similar to *BRCA1*. Taking into account the enrichment as well as the size of our case and control series, we had very good power (90%) to identify a new high-impact BCSG conferring a *BRCA1*-like breast cancer risk (OR = 10.6) with population combined mutation frequency ($MAF_{combined}$) down to (i) 0.0043% (or 1 in 11,628) in the targeted DNA repair gene analysis or (ii) $MAF_{combined}$ of 0.0056% (or 1 in 8945) in the WES analysis. For a new gene of *PALB2*-like (OR=5) or *ATM*-like (OR=2.1) breast cancer risk, we had 90% power to identify a gene of $MAF_{combined}$ down to (i) 0.026% (1 in 1946) or 0.37% (1 in 135) respectively in the DNA repair analysis and (ii) 0.034% (1 in 1475) and 0.5% (1 in 100) in the WES analysis (Figure 4, Table 1).

These estimates are likely to be inherently conservative, for although we have accounted for bilateral disease and the first affected FDR in our series where present, we have not accounted for the additional family members with breast cancer, young age of cases, and concomitant occurrence of ovarian cancer[20].

Power calculations representing v***ariability in population MAFcombined***, the largest breast WES study published to date, and expansion of the current study with increasing ***sample size***, ***are presented in Supplementary Tables S11***, ***S12,*** and S13, respectively.

Discussion

In our series of 2,135 enriched BRCA-negative female breast cancer cases, only 159 (7.4%) have a P/LP variant in an established BCSG. The vast majority (138/159, 87%) of these P/LP variants were in three genes: *PALB2*, *CHEK2* and *ATM*. As would be anticipated from our ascertainment framework, the contribution of P/LP variants in genes associated with pleiomorphic tumor syndromes was very modest (0.5% of cases). Furthermore, given that P/LP variants in *BARD1*, *RAD51C* and *RAD51D* are of relatively low frequency even in the triple-negative breast cancers with which they are associated, their contribution was predictably low in this series ascertained agnostic to histology (0.5% of cases). The frequency in cases of P/LP variants in other autosomal dominant CSGs (2.5%) was not elevated compared to controls (2.7%), suggesting firstly that these variants are largely incidental to etiology of these breast cancers and secondly that amongst other established CSGs there is minimal unexplained residual pleiotropy for breast cancer.

In our subsequent exploration of residual breast cancer genetic susceptibility attributable to rare variants, aside from 'rediscovery' of established BC susceptibility genes *PALB2*, *CHEK2* and *ATM*, no gene attained statistically significant signal of association at the respective thresholds for either the candidate-based approaches or the exome-wide approach. Furthermore, there was no evidence for association found in a replication series (UK Biobank) for top-ranking genes with nonsignificant signals of association.

To our knowledge, our study is the largest case-control analysis reported to date in enriched breast cancer cases, and better powered for discovery of a new rare high penetrance susceptibility gene than the largest reported case-control analysis to date of unselected breast cancers (7,859 cases/117,456 controls)[19] (Supplementary Table S2, Supplementary Table S12

Contextualization of our findings with previous studies

In the 1990s, linkage analysis in modest numbers of multi-case pedigrees enabled identification of *BRCA1* followed by *BRCA2*[31-34]. Subsequent linkage analyses in larger series of *BRCA1* and *BRCA2* negative breast cancer families yielded no reproducible signals[35]. Thereafter followed case-control mutational screening for genes functionally related to *BRCA1* and *BRCA2,* necessarily limited to modest numbers of samples on account of the low-throughput technologies available. Through these experiments, *PALB2, CHEK2* and *ATM* were identified as BCSGs, with the less reproducible associations for *BARD1*, *RAD51C* and *RAD51D* subsequently confirmed as exclusive to ER-negative tumours[25, 36-38].

It was widely anticipated that the advent of NGS would allow identification of the elusive additional BCSGs. Initial applications of WES typically involved case-only segregation-type analyses of multiple members of breast cancer families (Supplementary Table S2). With wider public availability of control data, NGS breast cancer gene discovery studies progressed onto case/control analyses, with the vast majority focused on targeted sets of DNA repair genes selected on the basis of biological candidacy [39-41]. Few case-control WES analyses for breast cancer have been reported. In 2021 Wang al. reported the largest WES case-control analysis in

breast cancer to date comprising 7,859 unselected breast cancer cases compared to 117,456 controls in UKBiobank[19]. Overall, despite almost a decade of application of NGS via family-only, candidate and exome-wide case-control approaches, including the study presented here, none has yielded robust novel associations of genes with breast cancer susceptibility.

Power analyses from this study demonstrate that whilst entirely plausible that additional high penetrance BCSGs exist, (i) for genes of *BRCA1*-like risks, the $MAF_{combined}$ will be vanishingly low (ii) for any gene in the ~5-fold range of risk, the population mutational frequency is likely to be less than half that of *PALB2* (i.e. less than 1 in 1500 in the population). These sobering power analyses contextualize the failure of those modest-sized early linkage studies to identify PALB2 and of the WES familial segregation studies to yield reproducible findings.

Conversely, our study was substantially less well powered for genes of more modest effect sizes, for which the use of enriched cases also offers less benefit (Table 1). For truncating variants in *CHEK2,* large genetic epidemiologic studies have demonstrated OR of 2.5 and population $MAF_{combined}$ of 0.31% (1 in 163)[25]. Our power to detect a gene of this risk/$MAF_{combined}$ profile was 95%, consistent with *CHEK2* coming out in our analysis as statistically significant for Class I. However, for *ATM* we only had 12% power for detection of class I variants as the penetrance is lower (OR=2.1) and the population mutational frequency is half that of CHEK2 ($MAF_{combined}$=0.14% (1 in 350)). Thus, it is entirely plausible that numerous additional moderate penetrance BCSGs exist of impact profiles (penetrance x mutational frequency) akin to that of

ATM and up to that of *CHEK2*, for which discovery studies to date (including ours) have been underpowered and/or not included in previously studied gene sets.

High penetrance genes are widely agreed to offer high clinical utility. Even recognizing variation in background polygenic contribution, the breast cancer risks are deemed sufficiently high to dichotomize management within families, with aggressive interventions such as risk-reducing mastectomies justified in family members carrying the mutation and restoration to near-population risk for those without. Contention persists regarding the utility for risk stratification offered by genes of more moderate penetrance, especially when conferring only sub-type specific risks and/or lacking recurrent founder mutations[5].

Limitations of study

Data regarding receptor status and/or histopathology was only available for a subset of patients, thus limiting power for histologically driven subtype analyses. The gnomAD series is only available as summary level count data; as such, equivalently stringent sample and variant level QC for the case data were required to address the possibility of inflation due to systematic genotyping error and potential differences in depth of coverage. This also meant we were unable to perform robust analyses for recessive and co-dominant susceptibility.

Another limitation was definitions by which missense variants were included as pathogenic/deleterious. For the CSG analysis, we only included only missense variants

classified on ClinVar as pathogenic/likely pathogenic with at least one star. Thus, any variants assigned on ClinVar as having 'Conflicting interpretations' were excluded. Likewise, for the DNA repair/exome-wide analyses, application of a threshold of REVEL > 0.7 was inevitably an imperfect filter, likely resulting both in inclusion of neutral variants and exclusion of deleterious variants.

Additionally, we did not include exon-level (or greater) deletions and duplications in our analyses. Such variants (which are inherently more difficult to detect from short read NGS data) are thought to comprise a non-trivial fraction of disease-associated germline alterations in cancer susceptibility genes. In the case of BRCA1 and BRCA2, they are thought to account for somewhere between 10-20% of all alterations in those genes. Thus, any putative novel breast cancer gene where structural variation is the predominant mutation type would be missed by our analyses. Our power analyses, therefore, do not account for any gain in power that would be afforded by the inclusion of such variants.

Future Studies

Future analysis of 15,000 similarly enriched breast cancer cases against 400,000 ethnicity-matched controls would be fully powered (> 99%) to identify all genes that might exist down to an *ATM*-like profile of risk and mutational frequency, as well as all genes of a *PALB2*-like risk (OR = 5), down to a population $MAF_{combined}$ of 0.0075% (1 in 6702) (Supplementary Table S13).

Conclusion

This is the largest study to date of enriched *BRCA*-negative breast cancer probands, in which 7.4% were found to carry a BCSG pathogenic mutation. With appropriate correction for multiple tests, aside from 'rediscovery' of known BCSGs, neither the targeted DNA repair gene analysis nor the WES approach yielded robust additional BCSGs. Outstanding BCSGs of high penetrance (OR>5) must be extremely rare. However, many genes up to a *CHEK2*/*ATM*-like risk may exist. Whilst a WES study involving five- to ten-fold more enriched breast cancer samples is feasible and would allow comprehensive survey of all such genes, there is not unanimous agreement regarding utility of clinical testing for genes that are so rare or of such modest penetrance.

Figure Captions

Figure 1. Study Overview. Overview of case and control samples used, along with generation and integration of datasets and analyses performed. BC, breast cancer; dmg. missense, damaging missense variant; HC PTV, high confidence protein truncating variant; QC, quality control; TCGA, The Cancer Genome Atlas.

F**igure 2. Distribution of case phenotypic features.** Distribution of phenotypes across 2,135 probands with invasive breast cancer where 'Family history' denotes at least one first degree or two second degree relatives affected with breast cancer, 'Ovarian' denotes the proband also being diagnosed with invasive ovarian cancer, 'Young onset' denotes the proband's first cancer being diagnosed at age <30, and 'Bilateral' denotes a diagnosis of bilateral breast cancer.

Figure 3: Frequency of deleterious variants in cases vs controls. Frequency of rare variants (MAF < 0.5%), collapsed by gene, comparing cases to controls for (a) high confidence protein truncating variants (HC-PTVs, as defined by LOFTEE) plus pathogenic/likely pathogenic variants (defined by ≥ one star review status ClinVar entries) in 86 established autosomal dominant cancer susceptibility genes; (b) HC-PTVs plus damaging missense with REVEL score ≥ 0.7 in 276 genes involved in DNA repair (c) HC-PTVs plus damaging missense exome-wide. The left panel in each

sub-figure shows the percentage of cases/controls and the right panel shows the corresponding odds ratio. Genes are ordered by *P* value. For each analysis, only the top 20 associations have been shown and the comparison is also shown for all genes, all breast-cancer susceptibility genes (BCSGs) and all non-BCSGs.

Figure 4: Power calculations in enriched vs unenriched breast cancer cases. The line plot demonstrates the power afforded by the current study of 2,135 genetically enriched breast cancer cases (solid lines) vs an unenriched (unselected) series of the same size (dashed lines) at exome-wide significance ($P < 1.3 \times 10^{-06}$). The histogram plot shows the number of genes with the corresponding control MAFcombined for each of the two variant class tests. OR, odds ratio.

Acknowledgement Section

Author Contributions

Loveday, Garrett and Turnbull contributed equally to this work. Loveday had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Loveday & Turnbull.

Acquisition of samples: Adlard, Barwell, Berg, Brady, Brewer, Chapman, Cook, Davidson, Donaldson, Douglas, Greenhalgh, Henderson, Izatt, Kumar, Lalloo, Miedzybrodzka, Morrison, Paterson, Porteous, Rogers, Walker, Eccles, Evans.

Analysis and interpretation of data: Loveday, Garrett, Law, Houlston, Turnbull.

Drafting of the manuscript: Loveday, Garrett, Turnbull & Houlston.

Critical revision of the manuscript for important intellectual content: Loveday, Garrett, Turnbull, Houlston, Miedzybrodzka, Hanson, Evans, Snape.

Statistical analysis: Loveday.

Obtained funding: Turnbull, Houlston.

Administrative, technical, or material support: Hanks, Poyastro-Pearson.

Study supervision: Turnbull.

Additional Contributions

We thank all the subjects and families that participated in the research. We thank those at the ICR, past and present, for their assistance in patient recruitment, sample management and management of the sequencing facility. We are grateful to all the clinicians and counsellors in The Breast and Ovarian Cancer Susceptibility Collaboration who have contributed to the recruitment and collection of samples. The full list of contributors is provided in the Appendix.

Conflict of Interest Disclosures

The authors have no conflicts of interest to disclose.

Sources of funding and support

This work was supported by Cancer Research UK [grants numbers C8620/A8372, C8620/A8857]; the Institute of Cancer Research [no applicable grant number]; and the National Health Service funding to the Royal Marsden / Institute of Cancer Research National Institute of Health Research Specialist Biomedical Research Centre for Cancer [no applicable grant numbers].

Role of funder(s)/sponsor(s)

The funder has not had any role in the analysis or interpretation of the data from the study. The funder did not have any role in the preparation, review, or approval of the manuscript.

References

1. Cancer Research UK Cancer Incidence. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero

2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. May 2021;71(3):209-249. doi:10.3322/caac.21660

3. Lynch HT, Krush AJ, Lemon HM, Kaplan AR, Condit PT, Bottomley RH. Tumor variation in families with breast cancer. JAMA. Dec 25 1972;222(13):1631-5.

4. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. Oct 27 2001;358(9291):1389-99. doi:10.1016/s0140-6736(01)06524-2

5. Turnbull C, Sud A, Houlston RS. Cancer genetics, precision prevention and a call to action. Nat Genet. Sep 2018;50(9):1212-1218. doi:10.1038/s41588-018-0202-0

6. Lilyquist J, Ruddy KJ, Vachon CM, Couch FJ. Common Genetic Variation and Breast Cancer Risk-Past, Present, and Future. Cancer Epidemiol Biomarkers Prev. Apr 2018;27(4):380-394. doi:10.1158/1055-9965.EPI-17-1144

7. Bahcall O. Common variation and heritability estimates for breast, ovarian and prostate cancers. Nature Genetics. 2013/03/20 2013;doi:10.1038/ngicogs.1

8. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. Nov 2 2017;551(7678):92-94. doi:10.1038/nature24284

9. Snape K, Ruark E, Tarpey P, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. Breast Cancer Res Treat. Jul 2012;134(1):429-33. doi:10.1007/s10549-012-2057-x

10. Guo MH, Plummer L, Chan YM, Hirschhorn JN, Lippincott MF. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. Am J Hum Genet. Oct 4 2018;103(4):522-534. doi:10.1016/j.ajhg.2018.08.016

11. Rahman N. Realizing the promise of cancer predisposition genes. Nature. Jan 16 2014;505(7483):302-8. doi:10.1038/nature12981

12. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. Nov 2018;18(11):696-705. doi:10.1038/s41568-018-0060-1

13. Knijnenburg TA, Vockley JG, Chambwe N, et al. Genomic and molecular characterization of preterm birth. Proc Natl Acad Sci U S A. Mar 19 2019;116(12):5819-5827. doi:10.1073/pnas.1716314116

14. Chakravarty D, Gao J, Phillips SM, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol. Jul 2017;2017doi:10.1200/PO.17.00011

15. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. May 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

16. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and

Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics. May 2015;17(5):405-24. doi:10.1038/gim.2015.30

17. Cubuk C, Garrett A, Choi S, et al. Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. Genet Med. Jul 6 2021;doi:10.1038/s41436-021-01265-z

18. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. Oct 6 2016;99(4):877-885. doi:10.1016/j.ajhg.2016.08.016

19. Wang Q, Dhindsa RS, Carss K, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature. Sep 2021;597(7877):527-532. doi:10.1038/s41586-021-03855-y

20. Antoniou AC, Easton DF. Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol. Nov 2003;25(3):190-202. doi:10.1002/gepi.10261

21. Brewer HR, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Family history and risk of breast cancer: an analysis accounting for family structure. Breast Cancer Res Treat. Aug 2017;165(1):193-200. doi:10.1007/s10549-017-4325-2

22. Kheirelseid EA, Jumustafa H, Miller N, et al. Bilateral breast cancer: analysis of incidence, outcome, survival and disease characteristics. Breast Cancer Res Treat. Feb 2011;126(1):131-40. doi:10.1007/s10549-010-1057-y

23. Seal S, Thompson D, Renwick A, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet. Nov 2006;38(11):1239-41. doi:10.1038/ng1902

24. Roberts ME, Jackson SA, Susswein LR, et al. MSH6 and PMS2 germ-line pathogenic variants implicated in Lynch syndrome are associated with breast cancer. Genet Med. Oct 2018;20(10):1167-1174. doi:10.1038/gim.2017.254

25. Dorling L, Carvalho S, Allen J, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med. Feb 4 2021;384(5):428-439. doi:10.1056/NEJMoa1913948

26. Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature. Jan 17 2013;493(7432):406-10. doi:10.1038/nature11725

27. Kim B, Won D, Lee ST, Choi JR. Somatic mosaic truncating mutations of PPM1D in blood can result from expansion of a mutant clone under selective pressure of chemotherapy. PLoS One. 2019;14(6):e0217521. doi:10.1371/journal.pone.0217521

28. Pharoah PDP, Song H, Dicks E, et al. PPM1D Mosaic Truncating Variants in Ovarian Cancer Cases May Be Treatment-Related Somatic Mutations. J Natl Cancer Inst. Mar 2016;108(3)doi:10.1093/jnci/djv347

29. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. May 2000;25(1):25-9. doi:10.1038/75556

30. Gene Ontology C. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. Jan 8 2021;49(D1):D325-D334. doi:10.1093/nar/gkaa1113

31. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. Dec 21 1990;250(4988):1684-9. doi:10.1126/science.2270482

32. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science. Sep 30 1994;265(5181):2088-90. doi:10.1126/science.8091231

33. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature. Dec 21-28 1995;378(6559):789-92. doi:10.1038/378789a0

34. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. Oct 7 1994;266(5182):66-71. doi:10.1126/science.7545954

35. Smith P, McGuffog L, Easton DF, et al. A genome wide linkage search for breast cancer susceptibility genes. Genes Chromosomes Cancer. Jul 2006;45(7):646-55. doi:10.1002/gcc.20330

36. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. May 2002;31(1):55-9. doi:10.1038/ng879

37. Thompson D, Duedal S, Kirner J, et al. Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst. Jun 1 2005;97(11):813-22. doi:10.1093/jnci/dji141

38. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. Feb 2007;39(2):165-7. doi:10.1038/ng1959

39. Lu HM, Li S, Black MH, et al. Association of Breast and Ovarian Cancers With Predisposition Genes Identified by Large-Scale Sequencing. JAMA Oncol. Jan 1 2019;5(1):51-57. doi:10.1001/jamaoncol.2018.2956

40. Huang KL, Mashl RJ, Wu Y, et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. Apr 5 2018;173(2):355-370 e14. doi:10.1016/j.cell.2018.03.039

41. Lu C, Xie M, Wendl MC, et al. Patterns and functional implications of rare germline variants across 12 cancer types. Nat Commun. Dec 22 2015;6:10086. doi:10.1038/ncomms10086

Appendix

Breast and Ovarian Cancer Susceptibility Collaboration

The following individuals are part of Breast and Ovarian Cancer Susceptibility Collaboration:

A. Ardern-Jones[1], J. Adlard[2], M. Ahmed[3], G. Attard[4], K. Bailey[5], E. Bancroft[1], C. Bardsley[6], D. Barton[7], M.Bartlett[8], J. Barwell[9], L. Baxter[6], R. Belk[10], J. Berg[11], B. Bernhard[8], T. Bishop[2], L. Boyes[12], N. Bradshaw,[13], A.F. Brady[8], S. Brant[14], C. Brewer[14], G. Brice[4], G. Bromilow[14], C. Brooks[15], A. Bruce[16], B. Bulman[10], L. Burgess[12], J. Campbell[17], N. Canham[8], B. Castle[3], R. Cetnarskyj[17], C. Chapman[12], O. Claber[18], N. Coates,[19], T. Cole[12], A. Collins[3], J. Cook[6], S. Coulson[15], G. Crawford[3], D. Cruger[12], C. Cummings[8], L. D'Mello[1], R. Davidson[13], L. Day[3], B. Dell[3], C. Dolling[12], A. Donaldson19, H. Dorkins[8], F. Douglas[18], S. Downing[20], S. Drummond[17], C. Dubras[3], J. Dunlop[11], S. Durrell[21], D. Eccles[3], C. Eddy[4], M. Edwards[17], E. Edwards[3], J. Edwardson[22], R. Eeles[1], I. Ellis[15], F. Elmslie[4], G. Evans[10], B. Gibbons[22], C. Gardiner[3], N. Ghali[8], C. Giblin[10], S. Gibson[13], S. Goff[4], S. Goodman[14], D. Goudie[11], L. Greenhalgh[15], J. Grier[5], H. Gregory[22], S. Halliday[1], R. Hardy[7], C. Hartigan[9], T. Heaton[12], A. Henderson[18], C. Higgins[15], S. Hodgson,[4], T. Homfray[4], D. Horrigan[5], C. Houghton[15], R.S. Houlston[1], L. Hughes[20], V. Hunt[14], L. Irvine[13], L. Izatt[23], C. Jacobs[23], S. James[22], M. James19, L. Jeffers[7], I. Jobson[18], W. Jones[4], M.J. Kennedy[5], S. Kenwrick,[20], C. Kightley[15], C. Kirk[7], E. Kirk[5], E. Kivuva[14], K. Kohut[1], M. Kosicka-Slawinska[8], A. Kulkarni[8], A. Kumar[16], F. Lalloo[10], N. Lambord[14], C. Langman[23], P. Leonard[12], S. Levene[23], S. Locker[5], P. Logan[7], M. Longmuir[13], A. Lucassen[3], V. Lyus[8], A. Magee[7], A. Male[16], S. Mansour[4], D. McBride[3], E. McCann[5],

V. McConnell[7], M. McEntagart[4], C. McKeown[12], L. McLeish[11], D. McLeod[21], A. Melville[8], L. Mercer[6], C. Mercer[3], Z. Miedzybrodzka[22], A. Mitra[1], P. J. Morrison[7], V. Murday[4], A. Murray[5], K. Myhill[1], J. Myring[11], E. O'Hara[2], J. Paterson[20], P. Pearson[17], G. Pichert[23], K. Platt[5], M. Porteous[17], C. Pottinger[15], S. Price[21], L. Protheroe[5], S. Pugh[19], O. Quarrell[6], K. Randhawa[8], C. Riddick[5], L. Robertson[1], A. Robinson,[20], V. Roffey-Johnson[9], M. Rogers[5], S. Rose[8], S. Rowe[10], A. Schofield[22], N. Rahman[1], S. Saya[1], G. Scott[13], J. Scott[7], A. Searle[14], S. Shanley[1], S. Sharif[12], A. Shaw[23], J. Shaw[15], J. Shea-Simonds[3], L. Side[16], J. Sillibourne[3], K. Simon[16], S. Simpson[22], S. Slater[17], S. Smalley[3], K. Smith[19], L. Snadden[13], K. Snape[1], J. Soloway[16], Y. Stait[14], B. Stayner[21], M. Steel[17], C. Steel[11], H. Stewart[21], D. Stirling[17], M. Thomas[14], S. Thomas,[1], S. Tomkins[19], H. Turner[5], A. Vandersteen[8], E. Wakeling[8], F. Waldrup[20], L. Walker[21], C. Watt,[13], S. Watts[23], A. Webber[15], C. Whyte[17], J. Wiggins[1], E. Williams[16], L. Winchester[4]

Enriched female breast cancer cases
N = 2,430

BRCA1/2 negative, plus one or more features of genetic enrichment:

- Bilateral disease
- Early onset BC
- Family history BC
- Ovarian cancer

gnomAD population controls
N = 51,377

Whole exome sequencing

illumina
nf-core/sarek germline
gatk

Exclusions (N = 295):

- sample QC (N = 192)
- cryptic relatendess (N = 1)
- occult BRCA muts (N = 42)
- ancestry (N = 60)

Whole exome sequencing
Summarised variant counts from Non-Finnish European, non-cancer population controls (i.e. ethnicity matched, TCGA removed)

Variant set integration and calibration
2,135 cases vs 51,377 controls

- Hard filtering (e.g. excess het)
- Calibration for depth adjusted quality
- Coverage normalisation

Rare variant burden testing

HC PTVs + ClinVar P/LP → Cancer susceptibility genes (N = 86)

Cancer driver genes (N = 686) → Oncogenic missense

DNA repair genes (N = 276) → HC PTVs +/- dmg. missense

Exome-wide (N = 19,651) → HC PTVs +/- dmg. missense

Power analyses

- Current study
- Enriched vs unenriched
- Future studies

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

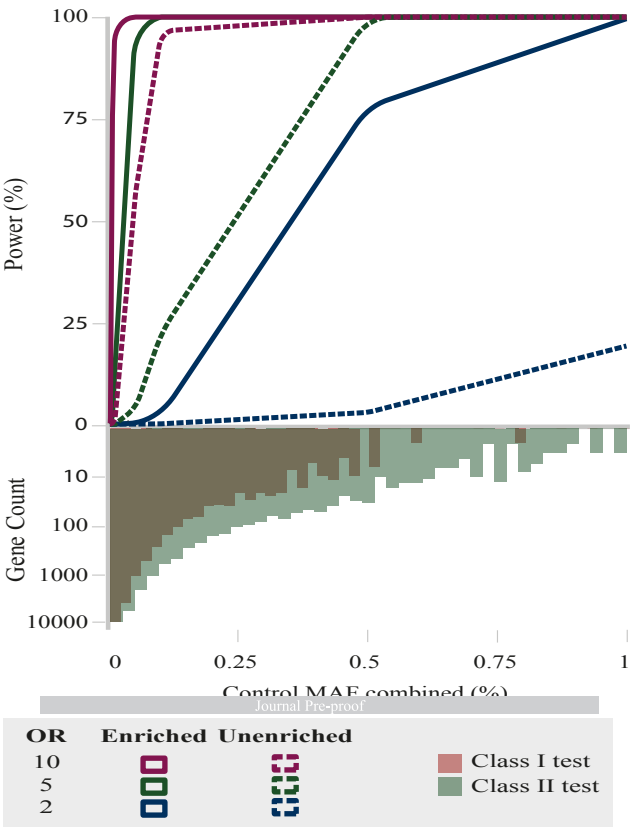
This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.



Enriched female breast cancer cases
N = 2,430

BRCA1/2 negative, plus one or more features of genetic enrichment:

- Bilateral disease
- Early onset BC
- Family history BC
- Ovarian cancer

gnomAD population controls
N = 51,377

Whole exome sequencing

illumina
nf-core/sarek germline
gatk

Exclusions (N = 295):

- sample QC (N = 192)
- cryptic relatendess (N = 1)
- occult BRCA muts (N = 42)
- ancestry (N = 60)

Whole exome sequencing
Summarised variant counts from Non-Finnish European, non-cancer population controls (i.e. ethnicity matched, TCGA removed)

Variant set integration and calibration
2,135 cases vs 51,377 controls

- Hard filtering (e.g. excess het)
- Calibration for depth adjusted quality
- Coverage normalisation

Rare variant burden testing

HC PTVs + ClinVar P/LP → Cancer susceptibility genes (N = 86)

Cancer driver genes (N = 686) → Oncogenic missense

DNA repair genes (N = 276) → HC PTVs +/- dmg. missense

Exome-wide (N = 19,651) → HC PTVs +/- dmg. missense

Power analyses

- Current study
- Enriched vs unenriched
- Future studies

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.

This is an Adobe® Illustrator® File that was saved without PDF Content.
To Place or open this file in other applications, it should be re-saved from Adobe Illustrator with the "Create PDF Compatible File" option turned on. This option is in the Illustrator Native Format Options dialog box, which appears when saving an Adobe Illustrator file using the Save As command.