BJC reports www.nature.com/bjcreports

ARTICLE OPEN

Check for updates

A comprehensive genomic framework for identifying genes predisposing to homologous recombination repair-deficient breast or ovarian cancer

José Camacho-Valenzuela[1,2,3], Thibaut Matis[4,5], Carla Roca[6,7], Jorge Luis Cuamatzi Flores[8], Nancy Hamel[2], Barbara Rivera[6,9], Simon Gravel[1], Paz Polak[10], Carla Daniela Robles-Espinoza[8,11] and William D. Foulkes[1,2,3,9]✉

BACKGROUND: Patients with clinical characteristics of increased cancer susceptibility without an identified genetic lesion are regularly seen in clinics. Association studies and matched normal/tumour sequencing have advanced the discovery of Cancer Susceptibility Genes (CSGs), with limitations when used independently. We reasoned that combining these strategies alongside mutational signatures and clinical data could improve CSGs identification.
METHODS: Using breast and ovarian cancer exome data from The Cancer Genome Atlas (TCGA-BRCA and TCGA-OV), we developed a genomic framework that evaluates exome-wide associations of Germline Pathogenic Variants (GPVs) harbouring second hits with Homologous Recombination Repair Deficiency (HRD) mutational signatures (HRDSig) to identify novel HRD-related CSGs. This is complemented by clinico-genomic analysis evaluating clinical and biological plausibility.
RESULTS: Our framework confirmed significant associations with HRDSig of *BRCA1/2* GPVs with second hits in both TCGA cohorts, validating its performance. *THBS4* also reached significance but only co-occurred with other HRD-related events in TCGA-BRCA. Borderline significance was also observed for *KIF13B* and *TESPA1*, also only in TCGA-BRCA. The clinico-genomics approach further identified *KIF13B* and *TESPA1*, as well as *RAD51B* and other Fanconi Anaemia pathway-related genes, including *FANCD2*, warranting further validation.
CONCLUSIONS: Our approach provides a framework for the identification of candidate HRD-related CSGs through combined statistical and clinico-genomics analyses. It is adaptable to other mutational signatures/cancer types and will be most effective when applied to larger and well-annotated datasets.

BJC Reports; https://doi.org/10.1038/s44276-026-00218-w

INTRODUCTION

It has been observed that approximately 7-8% of cancer cases overall are attributable to inherited factors such as Germline Pathogenic Variants (GPVs) in Cancer Susceptibility Genes (CSGs) [1, 2]. However, patients with clinical characteristics suggestive of increased cancer susceptibility that remain without identified GPVs in known CSGs are still regularly seen in the clinical setting, limiting options for genetic diagnosis, preventive strategies and treatment. This phenomenon, often known as "the missing heritability", could be explained in part by as-yet undiscovered GPVs in unrecognized cancer risk genes [3]. Incorporating these missing genes in cancer susceptibility clinical gene panels could improve clinical management of patients.

Different high-throughput sequencing methods have advanced our understanding of genetic susceptibility to cancer, including statistical association approaches [4–8] and analysis of matched normal/tumour sequencing data [9–11]. Each approach has specific advantages and limitations. Case-control studies are powerful to assess statistically significant gene associations across large groups of patients, but they do not consider the biological context of each tumour and generally limit their analysis to predefined gene lists [6, 7]. Meanwhile, in-depth study of the tumour mutational context of GPVs yields valuable biological insights [9, 10, 12–14], but such studies historically lacked statistical power due to small cohort size, though this is changing [11]. Combining both approaches, together with clinical data and mutational signatures into a unified framework, could theoretically enable the identification of CSGs that may have remained undetected.

The matched normal/tumour exome dataset of patients from The Cancer Genome Atlas (TCGA) is an extensively curated source of genomic and clinical data that can be exploited for such an integrated analytical approach. Furthermore, COSMIC (v3.4)

[1]Department of Human Genetics, McGill University, Montréal, QC, Canada. [2]Cancer Research Program, Centre for Translational Biology, The Research Institute of the McGill University Health Centre, Montréal, QC, Canada. [3]Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, QC, Canada. [4]Cancer Genetics Unit, Institut Bergonié, Bordeaux, France. [5]BRIC (BoRdeaux Institute of onCology), INSERM, Univ. Bordeaux, Bordeaux, France. [6]Bellvitge Biomedical Research Institute (IDIBELL), Avinguda de la Granvia de l'Hospitalet 199, L'hospitalet de Llobregat, Barcelona, Spain. [7]Doctoral Programme in Biomedicine, University of Barcelona (UB), Barcelona, Spain. [8]International Laboratory of Research of the Human Genome, National Autonomous University of Mexico, Querétaro, México. [9]Gerald Bronfman Department of Oncology, McGill University, Montréal, QC, Canada. [10]Quest Diagnostics, Secaucus, NJ, USA. [11]Experimental Cancer Genetics Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.
✉email: william.foulkes@mcgill.ca

Received: 18 October 2025 Revised: 18 February 2026 Accepted: 4 March 2026
Published online: 08 April 2026

Single-Base Substitution Signature 3 (SBS3), a mutational signature associated with Homologous Recombination Repair Deficiency (HRD), is highly prevalent in breast and ovarian cancer [12, 15–19]. In breast cancer, SBS3 and HRD are associated with the Triple-Negative Breast Cancer clinical subtype [13, 20], which overlaps in approximately 80% of cases with Basal-like Breast Cancer (BLBC) [21], one of the five intrinsic subtypes determined by the PAM50 classification [22, 23]. Similarly, in ovarian cancer, Sig3 and HRD are frequently observed in High-Grade Serous Ovarian Carcinoma (HGSOC) [17, 19].

In addition to SBS3, other mutational alterations have been associated with HRD, including single-base substitution and structural rearrangement signatures [18], and genomic instability features, such as telomeric allelic imbalance (TAI), large-scale transitions (LST), and loss of heterozygosity (LOH) [24], which have been observed in breast and ovarian cancer [10, 18, 25, 26]. Mechanistically, HRD typically results from biallelic inactivation of the well-known breast and ovarian cancer risk genes *BARD1, BRCA1/2, PALB2*, and *RAD51C/D* [9, 12–14, 18]. As such, the presence of HRD in a tumour could be used to infer the underlying biallelic inactivation of any gene (whether known or unknown) contributing to HRD [27].

In this study, we developed a comprehensive genomic framework that integrates association analysis, matched normal/tumour exome data, and clinical information from the TCGA breast (TCGA-BRCA) and ovarian (TCGA-OV) cancer cohorts. The framework aims to identify GPVs with somatic second hits in patients where the presence of HRD-associated mutational signatures in the tumour is not explained by genetic events in known HRD genes. Building on prior work [9, 10, 13, 14, 28, 29], we applied this framework to TCGA-BRCA and TCGA-OV independently, assessing within each cohort the statistical association of GPVs with somatic second hits in patients with and without HRD signatures, unrestricted to predefined gene sets. We integrated outputs from this dual statistical and patient-specific clinico-genomics approach with the aim of identifying new candidate HRD-related CSGs.

METHODS

Cohort data

We used exome data from the TCGA-BRCA and TCGA-OV cohorts (dbGAP Project #34072: phs000178.v11.p8). Blood-derived BAM files were downloaded from the Genomics Data Commons (GDC) platform. Somatic mutation data were obtained from the Multi-Center Mutation Calling in Multiple Cancers (MC3) project [30]. Copy Number Variant (CNVs) data were obtained from the GDC platform. Somatic promoter methylation data for *BRCA1* and *RAD51C* were obtained from a previous study [13], where they were analyzed for correlation with messenger RNA (mRNA) down-regulation. Clinical information was obtained from the cBioPortal. A summary of the datasets/sources is provided in Supplementary Table 1. Further details described in Supplementary Methods.

Sample inclusion criteria

SBS3 detection relies on assessing single-nucleotide somatic mutations in a tumour sample and analyzing their genomic context to identify specific patterns. To ensure reliable detection of SBS3, we established somatic mutation count thresholds. By performing bootstrap simulations in 14 positive controls downloaded from the COSMIC database (https://cancer.sanger.ac.uk/signatures/sbs/sbs3/), we determined a minimum threshold of 28 somatic mutations to detect SBS3 in all these positive controls. On the other hand, to eliminate hypermutated tumours, we considered 500 somatic mutations as the maximum threshold, as reported by Van Marcke et al., 2020 [29]. Selecting tumours with > 28 and < 500 somatic mutations from our cohorts resulted in 824 breast cancer and 378 ovarian cancer patients eligible for the subsequent analyses.

Bioinformatics pipeline for germline data

Blood-derived BAM files were processed using an in-house bioinformatics pipeline, and downstream exome filtering was performed according to predefined criteria (detailed in Supplementary Methods). Briefly, exome filtering included the following steps: (1) Variants flagged in gnomAD v2.1.1 as failing at least one of their quality controls were excluded; (2) Variants with >4 alternate reads, >20 total depth, and Variant Allele Fraction >0.1 were kept; (3) Exonic and splice site variants were retained, specifically missense, nonsense, frameshift INDELs, start-loss, and stop-loss; (4) Variants with population allele frequency < 0.01 in ExAC and/or gnomAD (or absent from both) were kept; (5) Variants classified in ClinVar as pathogenic or likely pathogenic were retained; (6) For variants annotated in ClinVar as conflicting, Variants of Uncertain Significance (VUS), or lacking classification, only missense, nonsense, splicing, and frameshift INDELs were further evaluated with stringent *in silico* predictions; (7) Genes with 3 or more variants in the same patient were excluded to reduce potential residual noise; (8) Olfactory receptor genes were also discarded due to false positive associations in cancer analyses [31]. The resulting final list of germline variants were considered as GPVs.

Second hit and mutational signature analyses

We determined the somatic second hit status of each GPV per patient in the cohort, either as (1) somatic point mutations/small INDELs or (2) as LOH through the loss of the wild-type allele (further details in Supplementary Methods). Of the 824 TCGA-BRCA cases, 30 had missing/unavailable CNV data and were therefore excluded. Similarly, of the 378 TCGA-OV cases, 12 were excluded on the same basis. This resulted in final cohorts of 794 TCGA-BRCA and 366 TCGA-OV patients that were included in all subsequent analyses.

For mutational signatures analysis, given that different classifiers to detect HRD-associated mutational patterns are optimally developed for breast and ovarian cancer [18, 32, 33], we used cohort-specific tools to define HRD status based on mutational signatures. For TCGA-BRCA tumours ($n = 794$), HRD status was inferred using the presence of SBS3 detected by SigMA [32], a computational algorithm specifically optimized for SBS3 detection in tumour samples with low mutation count. Further details are provided in the Supplementary Methods. For TCGA-OV tumours ($n = 366$), HRD status was defined using previously published HRDetect scores from Sztupinszki et al., 2021 [33] for the cases with available scores ($n = 353$). In that study, HRDetect, a Whole Genome Sequencing (WGS)-based HRD classifier that integrates multiple HRD-associated mutational alterations (e.g., SBS3, SBS8, rearrangement signatures 3 and 5, deletions with microhomology, TAI, LST, and LOH) [18], was trained to generate per-tumour HRD scores for TCGA-OV. Since HRDetect performance depends on tumour-specific training [34] and considering that their HRDetect scores were directly computed for TCGA-OV tumours [33], these scores are appropriate only for TCGA-OV. For clarity and consistency across both TCGA cohorts, we use the terms "HRD signature positive" (HRDSig+) and "HRD signature negative" (HRDSig-) to define tumours exhibiting HRD-related mutational patterns. Specifically, HRDSig status in TCGA-BRCA refers to SBS3 classification derived from SigMA, whereas in TCGA-OV it corresponds to HRDetect classification in TCGA-OV based on the scores generated by Sztupinszki et al., 2021 [33].

Unbiased association analysis approach

For TCGA-BRCA, we considered a tumour as HRDSig+ when having an SBS3 score >=0.5, derived from SigMA, as previously described [32]. Based on this threshold, we divided the breast cancer cohort ($n = 794$) into 2 subgroups: HRDSig+ group ($n = 141$) and HRDSig- group ($n = 653$). For TCGA-OV, tumours were classified as HRDSig+ when having an HRDetect score >=0.7, consistent with prior studies [18, 33]. Based on this cutoff, the ovarian cancer cohort ($n = 353$) was divided into 2 subgroups: HRDSig+ group ($n = 188$) and HRDSig- ($n = 165$). For each TCGA cohort, we calculated the frequency of carriers of GPVs with second hits in each gene of the HRDSig+ group compared to the frequency of carriers of GPVs with second hits in the HRDSig- group in the same gene, as shown in Fig. 1. We followed an unbiased exome-wide approach in both cohorts, where we did not restrict the analysis to any gene panel. In addition, to better detect signals from potentially new genes beyond the well-known HRD genes, we also performed a sub-analysis excluding patients carrying GPVs with second hits in *BARD1, BRCA1, BRCA2, PALB2, RAD51C,* and *RAD51D.*

Statistical analysis

We performed comparisons for the unbiased association analysis by evaluating, in a case-only design, whether the frequency of carriers of GPVs with second hits in each gene was enriched in HRDSig+ patients compared to HRDSig- cases within each TCGA cohort. We used Barnard's exact test

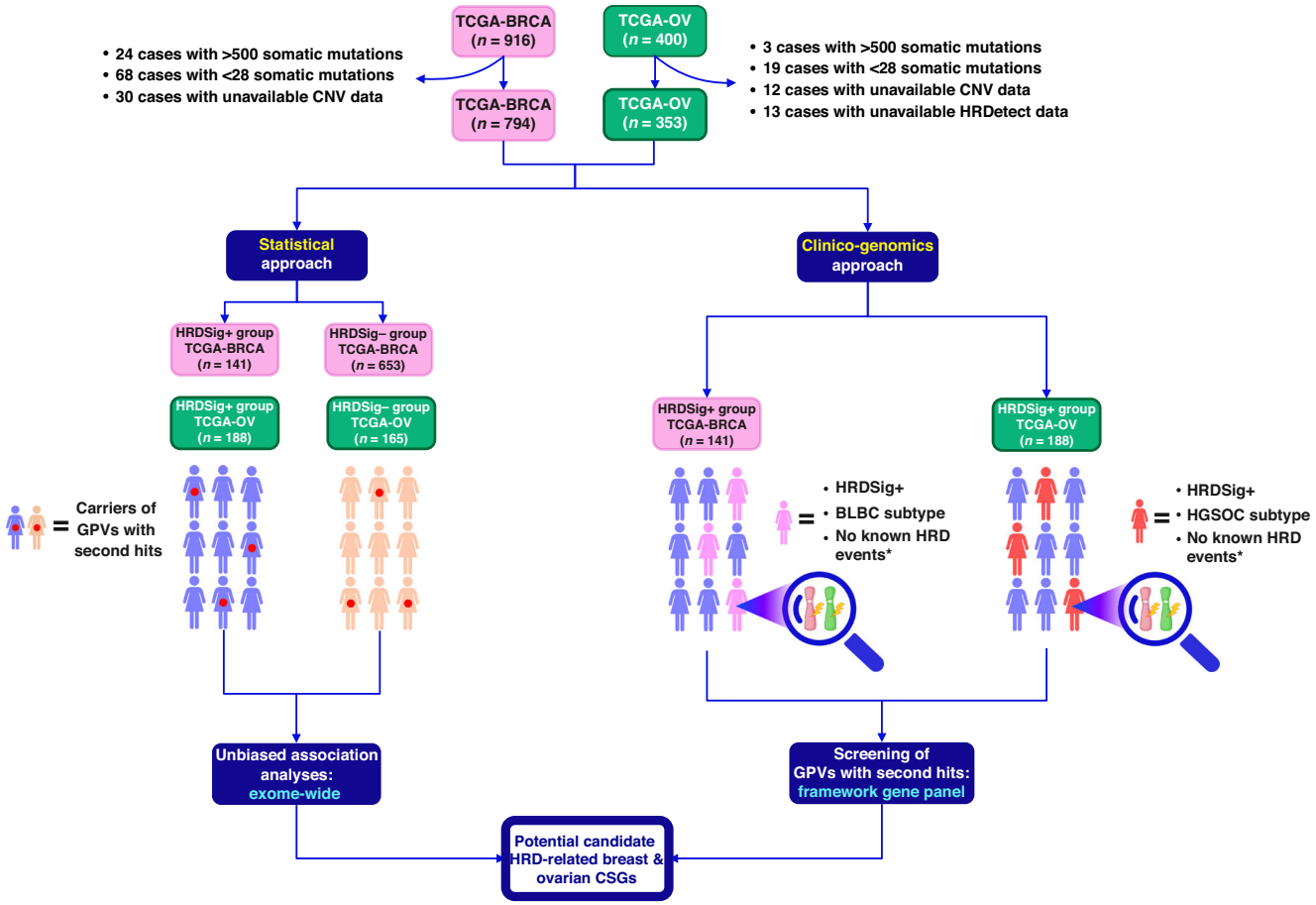

Fig. 1 Schematic overview of the comprehensive genomic framework. Unbiased association analyses are shown on the left. The clinico-genomics approach is shown on the right. Key: *No known HRD events: (1) GPVs with second hits and somatic homozygous deletion in the 6 well-known HRD-associated genes *BARD1, BRCA1/2, PALB2,* and *RAD51C/D*; (2) Somatic promoter methylation in *BRCA1* or *RAD51C*. Figure created with BioRender.com. TCGA-BRCA Breast cancer exome dataset of The Cancer Genome Atlas, TCGA-OV Ovarian cancer exome dataset of The Cancer Genome Atlas, CNV Copy Number Variants, HRDSig+ HRD signature positive, HRDSig- HRD signature negative, GPVs Germline Pathogenic Variants, HRD Homologous Recombination Repair Deficiency, BLBC Basal-like Breast Cancer, HGSOC High-Grade Serous Ovarian Carcinoma, CSGs Cancer Susceptibility Genes. File: Fig. 1.pdf.

(two-sided, $p < 0.05$), which does not condition on fixed margins for the evaluation of differences between 2 proportions analyzed in 2 × 2 contingency tables [35]. Such properties make this test particularly well-suited for small sample size datasets, conferring greater statistical power to evaluate associations in the context of low event counts.

Odds ratios and 95% Confidence Intervals (CIs) were calculated for each gene assessed with the Wald test. Here, the OR is measuring the magnitude of the association between the gene in question (harbouring a GPV with a second hit) with breast cancer that is HRDSig+ compared with breast cancer that is HRDSig-, not the overall risk of breast cancer. The same interpretation applied to TCGA-OV, defining the comparison between HRDSig+ and HRDSig- ovarian cancer. For genes with no observed events in the HRDSig- groups, the Haldane-Anscombe correction (adding 0.5 to genes with zero counts) was applied. Multiple-hypothesis testing was mitigated with the Benjamini-Hochberg method. Due to the relatively small sample size of our cohorts, a large number of genes had very low counts of GPVs with second hits (1 or 2 events), specifically 3626 out of 4369 (83%) from TCGA-BRCA, and 2931 out of 3297 (88%) from TCGA-OV, limiting the ability to estimate associations. Therefore, we only considered genes harbouring GPVs with second hits in at least five different patients as a minimum threshold to increase statistical power and reduce the risk of false associations driven by a very low count of events.

Clinico-genomics approach

To complement the unbiased association analyses, we incorporated biological principles (e.g., Knudson's two-hit hypothesis) into the analytical framework in an attempt to reconstruct the most likely tumorigenesis path for each tumour. We integrated germline, tumour and clinical data of all the individual patients from the HRDSig+ group from TCGA-BRCA ($n = 141$) and TCGA-OV ($n = 188$). We annotated validated types of genetic events that can explain these phenotypes on each tumour: both GPVs with second hits and somatic homozygous deletion in the six well-known HRD-associated genes *BARD1, BRCA1/2, PALB2* and *RAD51C/D*, as well as somatic promoter methylation in *BRCA1* or *RAD51C*. The rationale is that GPVs that co-exist with other genetic events already known to be linked to HRDSig+ status in tumours are less likely to be viable candidates than GPVs in patients where no other putative drivers of HRDSig+ status are observed.

We first looked for HRDSig+ patients with BLBC in TCGA-BRCA and HRDSig+ patients with HGSOC in TCGA-OV but lacking any annotated HRD-associated events that could explain the presence of these phenotypes. In these patients, we then identified GPVs that have second hits in the tumour from an expanded cancer susceptibility gene panel (hereafter referred to as 'framework gene panel'). This gene list has been curated in-house and created by merging clinical-grade panels for cancer susceptibility offered by different diagnostic laboratories, and also includes genes evaluated in two of the largest case-control studies in hereditary breast cancer [6, 7], as well as a broad list of DNA-repair genes (Supplementary Data 1). The inclusion of DNA-repair genes is based on the hypothesis that, in addition to *BARD1, BRCA1/2, PALB2* and *RAD51C/D*, other Homologous Recombination Repair (HRR) genes currently not known to be related to SBS3 (or genes from other DNA-repair pathways) may also contribute to HRD. We also included the genes reaching statistical significance in the exome-wide unbiased association analyses.

RESULTS

Identifying candidate CSGs using unbiased association analyses

In Fig. 1, we illustrate a schematic overview of the comprehensive genomic framework proposed in our study. The unbiased association analysis compares the per gene frequency of carriers of GPVs with second hits between HRDSig+ and HRDSig- patients in TCGA-BRCA and TCGA-OV. In a first global association analysis (without excluding carriers of GPVs with second hits in the well-established HRD genes), 221 genes were assessed in TCGA-BRCA and 90 genes in TCGA-OV. We observed in each cohort a significant association ($p < 0.05$) of HRDSig+ with the presence of GPVs with second hits in the proof of principle genes *BRCA1* and *BRCA2*. In TCGA-BRCA, *BRCA1* showed the highest odds ratio (OR = 38.7 [95% CI 8.7–171.5], adjusted *p*-value 1.8×10^{-8}), followed by *BRCA2* (OR = 30.2 [95% CI 6.6–136.9], adjusted p-value 2.06×10^{-6}), as illustrated in Fig. 2a and Supplementary Data 2A. Similarly, in TCGA-OV, GPVs with second hits in *BRCA1* (OR = 11.9 [95% CI 3.5–39.6], adjusted *p*-value 5.3×10^{-5}) and *BRCA2* (OR = 11.3 [95% CI 2.6–48.9], adjusted *p*-value 2.6×10^{-3}) reached the highest significant association with HRDSig+, as shown in Fig. 2b and Supplementary Data 2b. These observations were expected and validate the effectiveness of the framework. In this case-only design, the odds ratios indicate that *BRCA1/2* GPVs carriers with second hits are strongly enriched among HRDSig+ tumours (OR > 30 in TCGA-BRCA; OR > 10 in TCGA-OV).

In TCGA-BRCA, we also detected a significant association for *THBS4* (OR = 23.9 [95% CI 2.7–206.8], adjusted p-value 0.02), indicating enrichment in HRDSig+ breast cancer patients over HRDSig- cases. In TCGA-OV, no other gene reached statistical significance ($p < 0.05$) for association with presence of HRDSig+. Notably, some well-known HRD genes were not statistically tested in both cohorts because they did not meet the minimum threshold of harbouring GPVs with second hits in at least five different patients to qualify for statistical comparison. In TCGA-BRCA, this was the case for *BARD1* ($n = 1$), *PALB2* ($n = 2$), *RAD51C* ($n = 2$) and *RAD51D* ($n = 1$). In TCGA-OV, this was the case for *PALB2* ($n = 2$), *RAD51C* ($n = 1$) and *RAD51D* ($n = 4$). Interestingly, other breast and ovarian CSGs commonly tested clinically showed odds ratios < 1, although these associations were not statistically significant. In TCGA-BRCA, this was the case of *ATM* (OR = 0.2 [95% CI 0.03 – 2.1], adjusted p-value 0.54) and *CHEK2* (OR = 0.3 [95% CI 0.01–5.3], adjusted *p*-value 0.58), as described in Fig. 2a and Supplementary Data 2a. In TCGA-OV, this was the case of *BRIP1* (OR = 0.2 [95% CI 0.02–1.9], adjusted *p*-value 0.75), as shown in Fig. 2b and Supplementary Data 2b.

To better detect signals from potential novel genes beyond the already well-known HRD genes, we repeated the unbiased exome-wide association analysis after excluding all patients with GPVs with somatic second hits in the well-known HRD genes *BARD1, BRCA1, BRCA2, RAD51C, RAD51D* and *PALB2*. This resulted in 194 genes being analyzed in TCGA-BRCA and 54 genes in TCGA-OV. In the context of fewer gene comparisons tested, *THBS4* and three additional genes emerged as statistically significant. *THBS4, KIF13B* and *TESPA1* all had the same number of events in both HRDSig+ group ($n = 4$) and HRDSig- group ($n = 1$), (OR = 24.8 [95% CI 2.7–224.4], adjusted *p*-value 0.049), as shown in Fig. 3a and Supplementary Data 3a. For the fourth gene, *ATXN3* (OR = 4.3 [95% CI 2.04–9.3], adjusted *p*-value 0.04), most of the identified variants in both HRDSig groups were frameshift insertions clustered within a low complexity genomic region flagged in gnomAD as "lcr" (except for two variants, which were located only two and three base pairs outside this area, respectively). Such regions are characterized by repetitive sequences prone to alignment issues and sequencing artifacts. While we retained this gene in the association analyses to preserve an unbiased exome-wide approach, its biological interpretation is likely artifactual.

Despite reducing the number of statistical gene comparisons, no gene reached statistical significance ($p < 0.05$) in TCGA-OV after exclusions. For *BRIP1* in TCGA-OV (Fig. 3b and Supplementary Data 3b), as well as for *ATM* and *CHEK2* in TCGA-BRCA, the exclusion of carriers of GPVs with second hits in well-known HRD genes resulted in odds ratios that were as expected slightly closer to 1.0 (*ATM* (OR = 0.36 [95% CI 0.04–2.8], adjusted *p*-value 0.74); *CHEK2* (OR = 0.39 [95% CI 0.02–6.9], adjusted *p*-value 0.74); *BRIP1* (OR = 0.31 [95% CI 0.03–2.8], adjusted *p*-value 0.84).

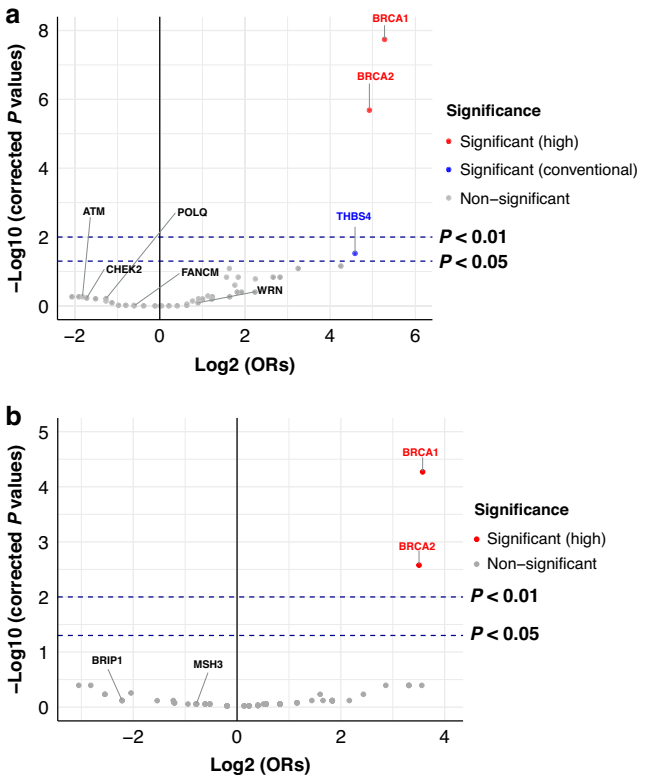

Fig. 2 Exome-wide association with HRDSig+ of GPVs with somatic second hits. Data are displayed as volcano plots of **a** TCGA-BRCA, and **b** TCGA-OV, including patients carrying GPVs with second hits in the well-known HRD genes *BARD1*, *BRCA1/2*, *PALB2*, and *RAD51C/D*. Barnard's exact test, two-sided. The X axis shows odds ratios (log2 scale). The Y axis shows adjusted *p*-values (-log10 scale). Horizontal dotted dark-blue lines indicate the conventional ($p < 0.05$) and strict ($p < 0.01$) significance thresholds. *P*-values shown were adjusted with the Benjamini-Hochberg method to mitigate multiple hypotheses testing. Red dots indicate genes reaching statistical significance at $p < 0.01$. The blue dot indicates the gene that reached statistical significance at $p < 0.05$. Gray dots indicate genes not reaching statistical significance. Non-significant genes labeled in the plot correspond to those included in our framework gene panel. ORs: Odds ratios.

Identifying candidate CSGs using a clinico-genomics approach

To complement the statistical approach, we applied a clinico-genomics strategy integrating clinical data (age at diagnosis, sex, TCGA-reported race, BLBC subtype for TCGA-BRCA and HGSOC subtype for TCGA-OV) with the corresponding germline and somatic mutational landscapes of HRDSig+ patients from both cohorts (141 in TCGA-BRCA and 188 in TCGA-OV, as illustrated in Supplementary Fig. 1a and b**)**. Specifically, we screened GPVs with second hits in genes from our in-house framework gene panel in HRDSig+ patients with BLBC (for TCGA-BRCA) and HGSOC (for TCGA-OV) that lacked any well-known HRD-associated events that could explain the presence of HRDSig+ in their tumours. For TCGA-BRCA, the framework panel was expanded to include the statistically significant non-*BRCA1/2* genes identified by our

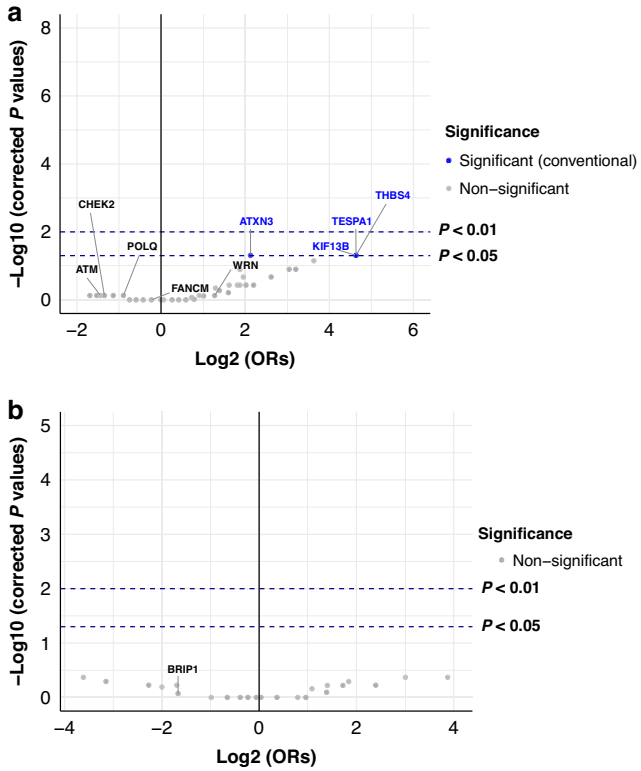

Fig. 3 Exome-wide association with HRDSig+ of GPVs with somatic second hits. Data are displayed as volcano plots of **a** TCGA-BRCA, and **b** TCGA-OV, excluding patients carrying GPVs with second hits in the well-known HRD genes *BARD1*, *BRCA1/2*, *PALB2*, and *RAD51C/D*. Barnard's exact test, two-sided. The X axis shows odds ratios (log2 scale). The Y axis shows adjusted p-values (-log10 scale). Horizontal dotted dark-blue lines indicate the conventional ($p < 0.05$) and strict ($p < 0.01$) significance thresholds. P-values shown were adjusted with the Benjamini-Hochberg method to mitigate multiple hypotheses testing. Blue dots indicate genes that reached statistical significance at $p < 0.05$. Gray dots indicate genes not reaching statistical significance. Non-significant genes labeled in the plot correspond to those included in our framework gene panel. ORs: Odds ratios.

unbiased association analyses, *KIF13B, TESPA1* and *THBS4*. For TCGA-OV, the framework panel was not expanded since no gene beyond *BRCA1/2* reached statistical significance.

Using this strategy in TCGA-BRCA, we identified 11 of 141 patients from the HRDSig+ group (7%) that carried 17 GPVs with second hits in 16 different genes (Fig. 4a and Supplementary Data 4a). Of these 11 patients, 6 cases harboured GPVs with second hits in 7 genes either involved in HRR or in an HRR-related pathway, including the HRR gene *RAD51B*, and the FA pathway-related genes *ATR, ERCC4, FAAP20, FAN1, FANCD2*, and *FANCM*. Of note, GPVs with second hits observed in *ATR, FAAP20, FAN1, FANCD2* and *RAD51B* were predicted to be truncating, including one frameshift deletion in *FAAP20*, two splicing in *FAN1* and *FANCD2*, and two nonsenses in *ATR* and *RAD51B*. In contrast, two missense GPVs with second hits were observed in *ERCC4* and *FANCM*. Clinico-genomics assessment highlighted additional CSGs and non-HRR DNA-repair genes (Fig. 4a and Supplementary Data 4a). In TCGA-OV, we identified 16 of 188 patients from the HRDSig+ group (8%) that carried 27 GPVs with second hits in 24 different genes (Fig. 4b and Supplementary Data 4b). Of these 16 cases, 7 patients carried GPVs with second hits in 7 genes either belonging to HRR or interacting with an HRR-related pathway. This was the case of the HRR genes *BRIP1, RAD54B* and *RAD54L*, as well as the FA pathway-related genes *ERCC1, FAAP24, FANCD2* and *UBE2T*. From these genes, GPVs with second hits observed in *BRIP1, ERCC1, FAAP24, FANCD2* and *UBE2T* were predicted to be truncating, including two frameshift INDELs in *BRIP1* and *FAAP24*, two splicing in *ERCC1* and *FANCD2*, and one nonsense in *UBE2T*, whereas missense GPVs with second hits were identified in *RAD54B* and *RAD54L*. The clinico-genomics approach flagged other CSGs and non-HRR DNA-repair genes (Fig. 4b and Supplementary Data 4b).

Distribution of HRD-related events across TCGA cohorts

To get a global view of the HRDSig+ group from both cohorts, we also examined HRDSig+ patients carrying the well-known HRD-related events predefined in our study that could explain HRDSig+ presence. For TCGA-BRCA, we identified 73/141 patients (51%) carrying a well-known HRD-related event, of whom 72/73 carried a single HRD-related event (Fig. 5a). For TCGA-OV, we identified 118/188 patients (62%) carrying a well-known HRD-related event (Fig. 5b).

BRCA1-related HRD events were the most frequent in both cohorts, with 38 cases in TCGA-BRCA (15/38 with *BRCA1* GPVs with second hits and 23/38 with *BRCA1* somatic promoter methylation), and 80 cases in TCGA-OV (34/80 with *BRCA1* GPVs with second hits, and 46/80 with *BRCA1* somatic promoter methylation). *BRCA2*-related HRD events were the second most frequent in both cohorts, with 15 cases in TCGA-BRCA (12/15 with *BRCA2* GPVs with second hits, and 3/15 with *BRCA2* somatic homozygous deletion) and 23 cases in TCGA-OV (23 with *BRCA2* GPVs with second hits). *RAD51C*-related HRD events were observed in both cohorts predominantly via promoter methylation, with 15 cases in TCGA-BRCA (2/15 with *RAD51C* GPVs with second hits, 12/15 with *RAD51C* somatic promoter methylation, and 1/15 with *RAD51C* somatic homozygous deletion) and 9 cases in TCGA-OV (1/9 with *RAD51C* GPV with second hit, and 8/9 with *RAD51C* somatic promoter methylation). Additional HRD-related events in *BARD1, PALB2* and *RAD51D* were less frequent. Only one case with a *BARD1* GPV with a second hit was observed in TCGA-BRCA, and none in TCGA-OV. *PALB2* GPVs with second hits were observed in both cohorts (2 cases in TCGA-BRCA and 2 cases in TCGA-OV). *RAD51D* GPVs with second hits were also observed in both cohorts, with 1 case in TCGA-BRCA and 4 cases in TCGA-OV. Finally, one case in TCGA-BRCA had co-occurrence of two different HRD-related events, specifically somatic promoter methylation in both *BRCA1* and *RAD51C*.

DISCUSSION

In this study, we developed and optimized a comprehensive genomic framework designed to investigate cancers where mutational signatures in the tumour can be leveraged to identify CSGs in the germline, with the goal to explain part of the missing heritability still observed in cancer genetics clinics. We built on established matched normal/tumour sequencing frameworks from previous studies [9, 10, 13, 29]. Our framework integrates a statistically unbiased association analysis of GPVs with second hits within the context of cancer-relevant mutational signatures. This is complemented by clinico-genomics evidence to provide biological and clinical plausibility for the statistical findings, while controlling for co-occurring events linked to the mutational signature under evaluation (here HRDSig+). This tailored approach allows us to link with higher confidence GPVs with second hits in candidate genes to the mutational signature under analysis, all within the context of the clinical subtype associated with such a signature.

We applied this framework to the well-curated breast and ovarian cohorts from TCGA, leveraging their link with HRD-associated mutational signatures [12, 13, 17–19]. Our framework is versatile and can be extended to different cancer types associated with mutational signatures (e.g., mismatch repair in colorectal cancer). Our results suggest that this framework will perform best when applied to larger cohorts that provide sufficient statistical power to capture rare events and to distinguish true associations from false positives.

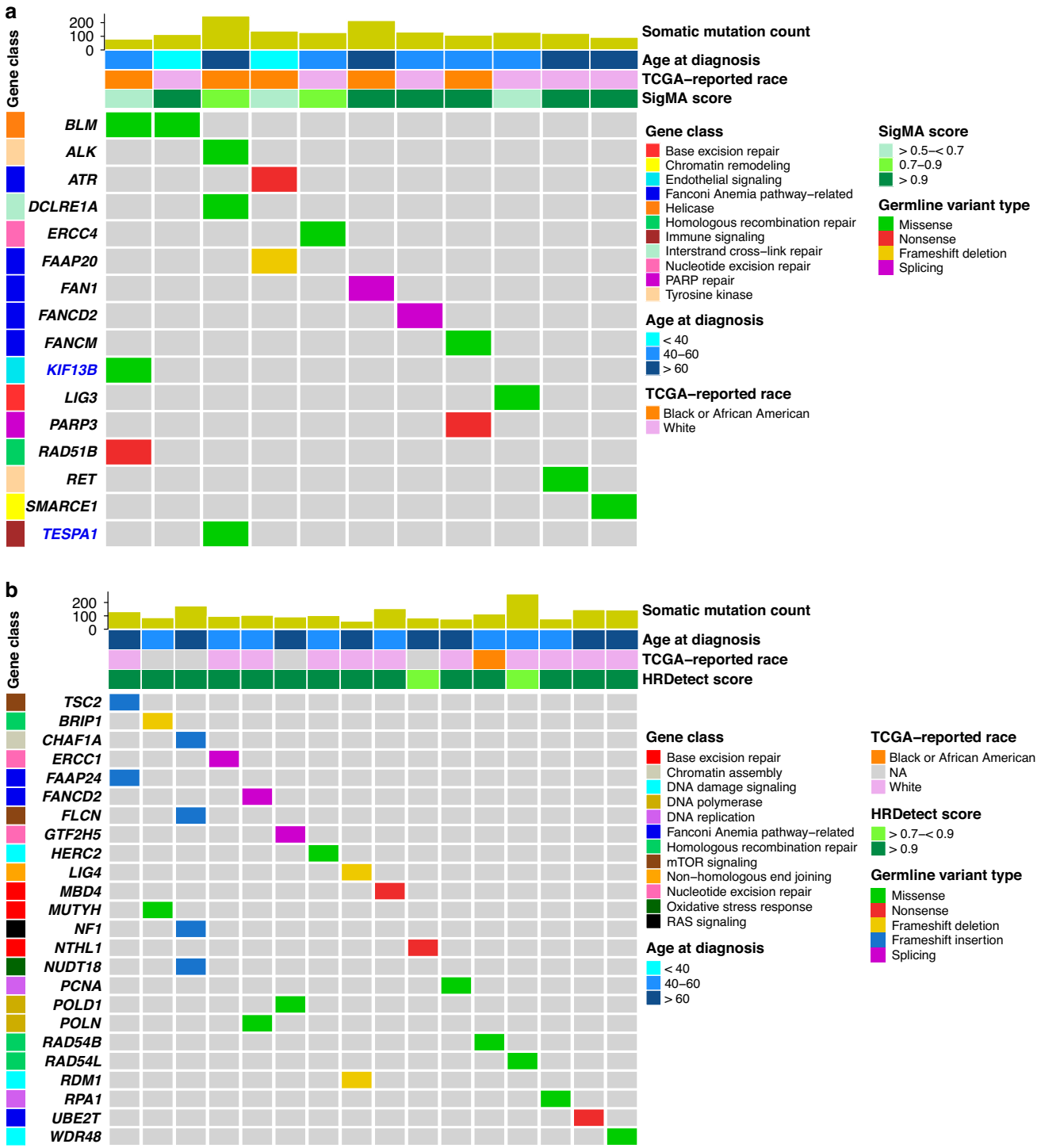

Fig. 4 Germline and somatic mutational landscape of HRDSig+ patients identified through the clinico-genomics approach. The oncoplots show **a** HRDSig+ patients ($n = 11/141$) with BLBC in the TCGA-BRCA cohort, and **b** HRDSig+ patients (n = 16/188) with HGSOC in the TCGA-OV cohort. All the GPVs illustrated harboured a second hit in the tumour, and no HRD-related genetic events (GPVs with second hits and somatic homozygous deletions in *BARD1, BRCA1/2, RAD51C/D*, and *PALB2*, as well as somatic promoter methylation in *BRCA1/RAD51C*) were observed in such patients. Gene labels in blue indicate genes that also reached statistical significance in the unbiased exome-wide association analyses ($p < 0.05$). TCGA-BRCA: Breast cancer exome dataset of The Cancer Genome Atlas, PARP: poly (ADP-ribose) repair, TCGA-OV: Ovarian cancer exome dataset of The Cancer Genome Atlas.

Statistically significant CSGs identified by association analyses

As expected, the unbiased exome-wide analysis identified *BRCA1/2* as the most significant susceptibility genes in both TCGA cohorts (Fig. 2a, b), reflecting their well-established association with HRD-associated mutational signatures [12, 13, 17–19]. These previously known associations also serve as proof-of-principle validation of our framework. Of note, the high odds ratios observed for *BRCA1/2* reflect the case-only design of our analysis (the odds of being HRDSig+ vs HRDSig-), not the case-control risk. For *BRCA1* in TCGA-BRCA, they also likely capture an HRDSig+ BLBC-specific signal,

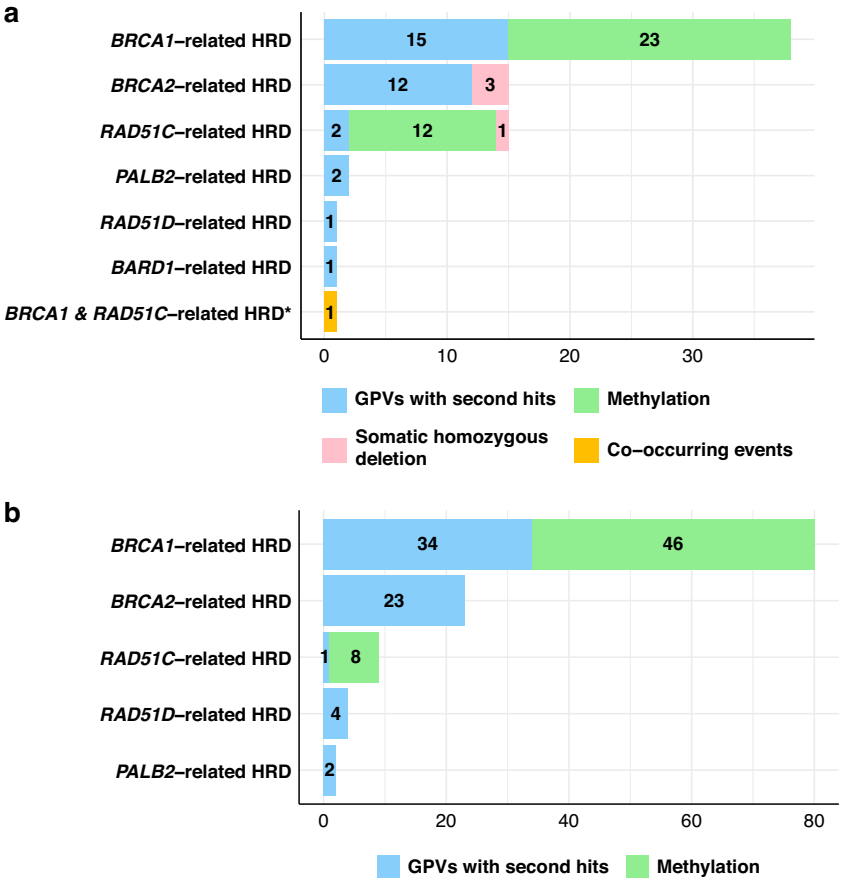

Fig. 5 Distribution of HRD-related events across TCGA cohorts. Barplots show **a** HRDSig+ patients ($n = 73/141$) in TCGA-BRCA and **b** HRDSig + patients (n = 118/188) in TCGA-OV with well-known HRD-related genetic events that potentially explains the presence of the mutational signature in question. Key: **BRCA1 & RAD51C*-related HRD: Patient with somatic promoter methylation in both *BRCA1* and *RAD51C*. HRD: Homologous Recombination Repair Deficiency, GPVs: Germline Pathogenic Variants.

since 14/15 carriers of *BRCA1* GPVs with second hits were BLBC. Similar odds ratios were reported in the CARRIERS study, where GPVs in *BRCA1* were strongly associated with TNBC [6]. In TCGA-OV, all carriers of *BRCA1* GPVs with second hits were HGSOC ($n = 34$), consistent with *BRCA1*'s known association with HGSOC [1, 36–38], but also reflecting the predominance of HGSOC in TCGA-OV. In TCGA-BRCA, *THBS4* also reached statistical significance in association analyses with and without exclusion of carriers of GPVs with second hits in well-known HRD genes. *THBS4* is an extracellular matrix gene previously suggested to be a tumour suppressor [39], implicated in breast cancer [40], and observed in hepatocellular and gastric cancers [41, 42]. However, clinico-genomics assessment showed that *THBS4* GPVs with second hits either co-occurred with other HRD-related events, more likely to explain the observed HRDSig +, or were found in non-BLBC tumours (Supplementary Fig. 1a).

KIF13B (involved in VEGFR2 trafficking and angiogenesis [43]) and *TESPA1* (involved in the regulation of antitumour immunity [44]) neared statistical significance ($p < 0.05$) in the global unbiased association analysis and reached significance when carriers of GPVs with second hits in known HRD-related genes were removed from the analysis. In contrast with *THBS4*, after evaluating them in their respective clinico-genomics context, GPVs with second hits in *KIF13B* and *TESPA1* were observed in HRDSig+ BLBC patients without known HRD-related events. For the *KIF13B* carrier, the co-occurrence of a *RAD51B* GPV with a second hit provides a more plausible explanation of the observed HRDSig+, given the role of *RAD51B* in HRR. In contrast, for the *TESPA1* carrier, no alternative genomic or promoter methylation event was found to explain the HRDSig+ status beyond the *TESPA1* GPV with the second hit. While its biological function was not previously linked to HRR [45], further investigation could be warranted. These potential associations were only observed in TCGA-BRCA, as no gene beyond *BRCA1/2* reached statistical significance in TCGA-OV, likely due to the small cohort size ($n = 353$) that limits statistical power.

CSGs identified through the clinico-genomics approach

Despite the limited cohort sizes, clinico-genomics assessment also revealed genes involved in HRR and the Fanconi Anaemia (FA) pathway, a DNA-repair mechanism related to HRR and linked to breast and ovarian cancer predisposition [46–48]. These genes carried GPVs with second hits, were observed in HRDSig+ patients lacking known HRD-related events and presenting HRD-associated clinical subtypes (BLBC in TCGA-BRCA and HGSOC in TCGA-OV). This included the HRR genes *RAD51B* in TCGA-BRCA and *RAD54B/L* in TCGA-OV. *RAD51B* and *RAD54B/L* promote DNA double-strand break repair via HRR [49–52]. *RAD51B*, a paralog of *RAD51C/D*, is further supported as a plausible HRD-related breast and ovarian CSG by an independent case-control study reporting a significant association with breast and ovarian cancer [53], and by recent WGS work identifying HRD tumours with biallelic *RAD51B* alterations [54]. *RAD54B/L* germline variants have been reported in breast [55–59] and ovarian [56, 58, 60–63] cancer cases and recently associated with HRD [64], with evidence of poly (ADP-ribose) polymerase (PARP) inhibitor sensitivity in ovarian cancer models with *RAD54B* alterations [65].

Convergence in both cohorts was observed for multiple FA pathway-related genes, including *ATR, FAAP20, FAN1, FANCD2*, and *FANCM* for TCGA-BRCA, and *FAAP24, FANCD2*, and *UBE2T* for TCGA-OV, all of which functionally interact with this pathway [48, 66]. Of note, *FANCD2* was identified in both TCGA cohorts. An independent case-control study found a significant association of *FANCD2* with hereditary breast cancer [67]. Likewise, germline *FANCD2* variants have been observed in ovarian cancer patients [68, 69] (including HGSOC cases [70]) and recently associated with HRD tumours [64]. In TCGA-BRCA, the 4 carriers of GPVs with second hits in *ATR, FAAP20, FAN1, FANCD2* and *FANCM* shared notable features: three out of four cases were diagnosed younger than 50 years of age, and three out of four were reported as Black or African American. GPVs in *ATR* [71, 72] and *FAN1* [56, 73] have been observed in hereditary breast cancer cases. In TCGA-OV, the three carriers of GPVs with second hits in *FAAP24, FANCD2*, and *UBE2T* were diagnosed older than 50 years of age and were reported as White. *FAAP20* has been found to play a strong role in HRR [74], whereas germline *FAAP24* variants have been observed in familial colorectal cancer [75], with evidence of PARP inhibitor sensitivity in prostate cancer models with *FAAP24* deficiency [76]. For *UBE2T*, also known as *FANCT*, biallelic germline variants have been previously reported to cause FA [77–79].

Additional DNA-repair genes with roles in the FA pathway were highlighted: *BRIP1* (also named *FANCJ* [80]) in TCGA-OV, as well as the nucleotide excision repair genes *ERCC4* (also named *FANCQ* [81]) in TCGA-BRCA, and *ERCC1* in TCGA-OV. Germline variants in *ERCC1* and *ERCC4* have been reported in breast and ovarian cancer [82–85]. *BRIP1* is a well-established ovarian CSG involved in HRR [36, 86–88] but not generally associated with HRD [12, 13]. In our study, the single carrier of a *BRIP1* GPV with a second hit also carries a GPV with a second hit in the base excision repair gene *MUTYH*, suggesting that any contribution of these genes to HRD, if any, may be the result of combined DNA-repair defects. Finally, while we observed a suggestive but non-significant ($p < 0.05$) negative association of *FANCM* with HRDSig+ in our statistical analyses, the clinico-genomics approach highlighted this gene in a Black or African American HRDSig+ case with BLBC that lacked any known HRD events (Fig. 4a). However, this patient also carried a nonsense *PARP3* GPV with a second hit, raising questions about the respective contribution of each gene, if any, to the HRD phenotype. In the HRDSig- control group, 4/7 *FANCM* carriers of GPVs with second hits were also BLBC, and 2/7 were reported as Black or African American (Supplementary Data 5). These findings are consistent with prior studies that link *FANCM* with BLBC [89–91], but they may also suggest that the contribution of *FANCM* to HRD may depend on additional gene-specific co-occurring variants and/or population-specific genetic factors. For instance, a recent study reported a strong association of *FANCM* with oestrogen receptor-negative breast cancer in a Hispanic/Latino cohort [92], an association not observed in two of the largest case-control studies conducted in predominantly European cohorts [6, 7].

Together, the clinico-genomics component of our framework demonstrates its potential to highlight biologically plausible candidate CSGs, including HRR and FA pathway-related genes. However, based on the TCGA cohorts only, their candidacy cannot be determined, as they could not be statistically tested due to low event counts. Rather than establishing definitive gene-disease association discoveries, these findings illustrate the framework's ability to flag potential candidates that warrant further assessment in larger datasets, and demonstrate the framework's versatility to be applied in different cancer types.

Comparison with previous studies of HRD and novel breast CSGs

Prior studies have combined matched normal/tumour sequencing data with SBS3 in predefined gene sets to investigate non-*BRCA1/2* familial breast cancer cases [9, 10, 14, 29], to directly associate SBS3 with candidate CSGs [13], or to compare mutational signatures in non-*BRCA1/2* breast tumours harbouring candidate gene alterations with mutational signatures from *BRCA1/2*-mutated tumours [28]. Our framework builds on these studies by requiring GPVs under study to have a somatic second hit to test their association with HRDSig+ exome-wide (beyond predefined gene sets) in TCGA-BRCA and TCGA-OV. In addition, it incorporates a layer of control for co-occurring events extended to all the known HRD genes, while leveraging BLBC and HGSOC clinical subtypes to guide the identification of candidate breast and ovarian CSGs. Together, these methodological refinements yield a framework that preserves and builds on the strengths of the previous approaches discussed above, while using additional stringent criteria to attribute a possible HRDSig+ contribution to candidate genes and variants.

Distribution of HRD-related events across TCGA cohorts

Among the 329 HRDSig+ patients from TCGA-BRCA ($n = 141$) and TCGA-OV ($n = 188$), somatic promoter methylation in *BRCA1* and *RAD51C* (69 and 20, respectively) was more frequent than carrying a GPV with second hits in these genes (49 and 3, respectively). In TCGA-BRCA, including patients with somatic homozygous deletion in *BRCA2* ($n = 3$) increased the number of *BRCA2*-related HRD events by 20% compared to *BRCA2* GPVs with second hits alone ($n = 12$). This indicates that germline testing alone may underestimate HRD prevalence in breast and ovarian cancer patients, with clinical implications with regard to PARP inhibitor eligibility.

The observed gene-specific distribution of HRD-related events in TCGA-BRCA is consistent with prior studies, including Polak et al. [13], who identified *BRCA1/2* and *RAD51C* as the primary SBS3 drivers in TCGA-BRCA, and the original TCGA-OV study, which showed a higher prevalence of *BRCA1* methylation compared to *BRCA1* germline variants in TCGA-OV [93]. Our framework builds on these previous studies by focusing exclusively on biallelic HRD events, restricting *BRCA1/RAD51C* promoter methylation analysis to HRDSig+ tumours, directly comparing these events to GPVs with second hits and somatic homozygous deletions (also in HRDSig+ tumours) and expanding the analysis to additional validated HRD genes (including *RAD51D*), all in the context of the breast and ovarian cancer TCGA cohorts. This analytical strategy is not limited to HRD. It can be adapted to annotate validated genomic events from other cancer-associated mutational signatures.

Framework limitations

Our framework's approach has limitations. First, the use of whole exome sequencing, while more accessible than WGS, excludes non-coding genetic variants possibly contributing to cancer risk and HRD. Second, since our framework evaluates only GPVs with second hits, it may overlook monoallelic variants potentially acting through haploinsufficiency. Third, the framework would require larger sample sizes to support or refute the candidacy of the genes of interest reported here as potential HRD-related CSGs. The relatively small sample size of the TCGA-BRCA and TCGA-OV datasets limited the statistical power to evaluate associations, requiring the adoption of a threshold for considering genes with GPVs with second hits in at least five different patients to minimize the risk of false associations, which excluded some genes from the analysis. In addition, although our framework was able to highlight biologically plausible candidate CSGs, the predominance of European-ancestry patients in TCGA restricted the discovery of potential ancestry-related GPVs [94]. Generating large and more ethnically diverse datasets with complete germline and tumour sequencing, methylation and copy number data that can unlock the full power of our framework will be resource intensive.

Finally, our analysis in TCGA-BRCA relies on SBS3 as the primary measure of HRD. Tumours with other HRD-associated mutational signatures and genomic instability features may not be captured

by SBS3 alone, but detection of these genomic alterations requires WGS, as exome data generally limits their identification. Consistent with this, WGS-based integrative approaches [18, 25, 54] have demonstrated the value of combining multiple genomic features for HRD detection, compared with individual metrics (e.g., genomic scars or *BRCA1/2* status) that capture only partial dimensions of the full spectrum of HRD-associated mutational alterations [95]. Future application of our framework to WGS datasets using WGS-based integrative approaches would be beneficial for a more robust classification of HRD.

CONCLUSIONS

In summary, we present a comprehensive genomic framework that integrates multiple types of germline/tumour genetic events and clinical information to identify novel CSGs and alleles. The versatility of our framework allows for the investigation of multiple mutational signatures and cancer types. When applied to the TCGA-BRCA and TCGA-OV cohorts, our framework highlighted convergence in both datasets of multiple candidate CSGs involved in HRR and the FA pathway, two biologically connected DNA-repair mechanisms. These results support previous data suggesting that *RAD51B* is an HRD-related breast and ovarian CSG. Among others, these genes deserve further study for validation, especially their inclusion in both larger and ancestrally diverse case-control and matched normal/tumour sequencing research efforts.

DATA AVAILABILITY

Germline sequencing data (blood-derived BAM files) analyzed in this study are available via the Genomic Data Commons (TCGA-BRCA, dbGaP Project #34072: phs000178.v11.p8). Somatic mutation data was obtained from the Multi-Center Mutation Calling in Multiple Cancers (MC3) project. Copy number variation data was downloaded from the Genomic Data Commons. Somatic promoter methylation data for *BRCA1* and *RAD51C* was obtained from a previous study [13]. Clinical data was downloaded from the cBioPortal. Further details and access links are provided in Supplementary Table 1 and Supplementary Methods. The code used to perform the statistical and clinico-genomics analyses from the genomic framework presented in this study is publicly available via GitHub (https://github.com/jcv333).

REFERENCES

1. Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. 2018;173:355–70.e14.
2. Yap TA, Ashok A, Stoll J, Mauer E, Nepomuceno VM, Blackwell KL, et al. Prevalence of germline findings among tumors from cancer types lacking hereditary testing guidelines. JAMA Netw Open. 2022;5:e2213070.
3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747–53.
4. Artomov M, Joseph V, Tiao G, Thomas T, Schrader K, Klein RJ, et al. Case-control analysis identifies shared properties of rare germline variation in cancer predisposing genes. Eur J Hum Genet. 2019;27:824–8.
5. Rowlands CF, Allen S, Balmaña J, Domchek SM, Evans DG, Hanson H, et al. Population-based germline breast cancer gene association studies and meta-analysis to inform wider mainstream testing. Ann Oncol. 2024;35:892–901.
6. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A population-based study of genes previously implicated in breast cancer. N Engl J Med. 2021;384:440–51.
7. Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, Wahlström C, et al. Breast cancer risk genes - association analysis in more than 113,000 women. N Engl J Med. 2021;384:428–39.
8. Wilcox N, Dumont M, González-Neira A, Carvalho S, Joly Beauparlant C, Crotti M, et al. Exome sequencing identifies breast cancer susceptibility genes and defines the contribution of coding variants to breast cancer risk. Nat Genet. 2023;55:1435–9.
9. Nones K, Johnson J, Newell F, Patch AM, Thorne H, Kazakoff SH, et al. Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers. Ann Oncol. 2019;30:1071–9.
10. Andersen LVB, Larsen MJ, Davies H, Degasperi A, Nielsen HR, Jensen LA, et al. Non-BRCA1/BRCA2 high-risk familial breast cancers are not associated with a high prevalence of BRCAness. Breast Cancer Res. 2023;25:69.
11. Sosinsky A, Ambrose J, Cross W, Turnbull C, Henderson S, Jones L, et al. Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 Genomes Cancer Programme. Nat Med. 2024;30:279–89.
12. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534:47–54.
13. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017;49:1476–86.
14. Matis TS, Zayed N, Labraki B, de Ladurantaye M, Matis TA, Camacho Valenzuela J, et al. Current gene panels account for nearly all homologous recombination repair-associated multiple-case breast cancer families. NPJ Breast Cancer. 2021;7:109.
15. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.
16. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.
17. Patch A-M, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole–genome characterization of chemoresistant ovarian cancer. Nature. 2015;521:489–94.
18. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23:517–25.
19. Koskela H, Li Y, Joutsiniemi T, Muranen T, Isoviita VM, Huhtinen K, et al. HRD related signature 3 predicts clinical outcome in advanced tubo-ovarian high-grade serous carcinoma. Gynecol Oncol. 2024;180:91–8.
20. Moore GM, Powell SN, Higginson DS, Khan AJ. Examining the prevalence of homologous recombination repair defects in ER+ breast cancers. Breast Cancer Res Treat. 2022;192:649–53.
21. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010;363:1938–48.
22. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–52.
23. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160–7.
24. Sztupinszki Z, Diossy M, Krzystanek M, Reiniger L, Csabai I, Favero F, et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. NPJ Breast Cancer. 2018;4:16.
25. Nguyen L, W. M. Martens J, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. Nat Commun. 2020;11:5584.
26. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous recombination Deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. Clin Cancer Res. 2016;22:3764–73.
27. Hughley R, Karlic R, Joshi H, Turnbull C, Foulkes WD, Polak P. Etiologic Index - A Case-Only Measure of BRCA1/2-Associated Cancer Risk. N Engl J Med. 2020;383:286–8.
28. Bodily WR, Shirts BH, Walsh T, Gulsuner S, King MC, Parker A, et al. Effects of germline and somatic events in candidate BRCA-like genes on breast-tumor signatures. PLoS One. 2020;15:e0239197.
29. Van Marcke C, Helaers R, De Leener A, Merhi A, Schoonjans CA, Ambroise J, et al. Tumor sequencing is useful to refine the analysis of germline variants in unexplained high-risk breast cancer families. Breast Cancer Res. 2020;22:36.
30. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. 2018;6:271–81.e7.
31. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499:214–8.
32. Gulhan DC, Lee JJ, Melloni GEM, Cortés-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nat Genet. 2019;51:912–9.
33. Sztupinszki Z, Diossy M, Borcsok J, Prosz A, Cornelius N, Kjeldsen MK, et al. Comparative assessment of diagnostic homologous recombination deficiency-associated mutational signatures in ovarian cancer. Clin Cancer Res. 2021;27:5681–7.
34. Štancl P, Hamel N, Sigel KM, Foulkes WD, Karlić R, Polak P. The great majority of homologous recombination repair-deficient tumors are accounted for by established causes. Front Genet. 2022;ume 13:2022.
35. Mehrotra DV, Chan IS, Berger RL. A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. Biometrics. 2003;59:441–50.
36. Morgan RD, Burghel GJ, Flaum N, Schlecht H, Clamp AR, Hasan J, et al. Extended panel testing in ovarian cancer reveals BRIP1 as the third most important predisposition gene. Genet Med. 2024;26:101230.

37. Witjes VM, van Bommel MHD, Ligtenberg MJL, Vos JR, Mourits MJE, Ausems M, et al. Probability of detecting germline BRCA1/2 pathogenic variants in histological subtypes of ovarian carcinoma. A meta-analysis. Gynecol Oncol. 2022;164:221–30.
38. Kim SI, Lee M, Kim HS, Chung HH, Kim JW, Park NH, et al. Effect of BRCA mutational status on survival outcome in advanced-stage high-grade serous ovarian cancer. J Ovarian Res. 2019;12:40.
39. Greco SA, Chia J, Inglis KJ, Cozzi SJ, Ramsnes I, Buttenshaw RL, et al. Thrombospondin-4 is a putative tumour-suppressor gene in colorectal cancer that exhibits age-related methylation. BMC Cancer. 2010;10:494.
40. McCart Reed AE, Song S, Kutasovic JR, Reid LE, Valle JM, Vargas AC, et al. Thrombospondin-4 expression is activated during the stromal response to invasive breast cancer. Virchows Arch. 2013;463:535–45.
41. Kuroda K, Yashiro M, Sera T, Yamamoto Y, Kushitani Y, Sugimoto A, et al. The clinicopathological significance of Thrombospondin-4 expression in the tumor microenvironment of gastric cancer. PLoS One. 2019;14:e0224727.
42. Su F, Zhao J, Qin S, Wang R, Li Y, Wang Q, et al. Over-expression of Thrombospondin 4 correlates with loss of miR-142 and contributes to migration and vascular invasion of advanced hepatocellular carcinoma. Oncotarget. 2017;8:23277–88.
43. Waters SB, Dominguez JR, Cho HD, Sarich NA, Malik AB, Yamada KH KIF13B-mediated VEGFR2 trafficking is essential for vascular leakage and metastasis in vivo. Life Sci Alliance. 2022;5.
44. Yang R, Yang M, Wu Z, Liu B, Zheng M, Lu L, et al. Tespa1 deficiency reduces the antitumour immune response by decreasing CD8(+)T cell activity in a mouse Lewis lung cancer model. Int Immunopharmacol. 2023;124:110865.
45. Liang J, Lyu J, Zhao M, Li D, Zheng M, Fang Y, et al. Tespa1 regulates T cell receptor-induced calcium signals by recruiting inositol 1,4,5-trisphosphate receptors. Nat Commun. 2017;8:15732.
46. Patel KJ. Fanconi anemia and breast cancer susceptibility. Nat Genet. 2007;39:142–3.
47. Niraj J, Färkkilä A, D'Andrea AD. The Fanconi Anemia Pathway in Cancer. Annu Rev Cancer Biol. 2019;3:457–78.
48. Bogliolo M, Surrallés J. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. Curr Opin Genet Dev. 2015;33:32–40.
49. Golmard L, Castéra L, Krieger S, Moncoutier V, Abidallah K, Tenreiro H, et al. Contribution of germline deleterious variants in the RAD51 paralogs to breast and ovarian cancers. Eur J Hum Genet. 2017;25:1345–53.
50. Suwaki N, Klare K, Tarsounas M. RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. Semin Cell Dev Biol. 2011;22:898–905.
51. McAndrew EN, McManus KJ. The enigmatic oncogene and tumor suppressor-like properties of RAD54B: Insights into genome instability and cancer. Genes Chromosomes Cancer. 2017;56:513–23.
52. Selemenakis P, Sharma N, Uhrig ME, Katz J, Kwon Y, Sung P, et al. RAD51AP1 and RAD54L Can Underpin Two Distinct RAD51-Dependent Routes of DNA Damage Repair via Homologous Recombination. Front Cell Dev Biol. 2022;10:866601.
53. Setton J, Selenica P, Mukherjee S, Shah R, Pecorari I, McMillan B, et al. Germline RAD51B variants confer susceptibility to breast and ovarian cancers deficient in homologous recombination. NPJ Breast Cancer. 2021;7:135.
54. Al Assaad M, Hadi K, Levine MF, Guevara D, Patel M, Tranquille M, et al. Whole genome sequencing approach to assess homologous recombination deficiency in a pan-cancer cohort. Commun Med. 2026;6:5.
55. Shumilova S, Danishevich A, Nikolaev S, Krasnov G, Ikonnikova A, Isaeva D, et al. High- and moderate-risk variants among breast cancer patients and healthy donors enrolled in multigene panel testing in a population of Central Russia. Int J Mol Sci. 2024;25.
56. Felicio PS, Grasel RS, Campacci N, de, Paula AE, Galvão HCR, et al. Whole-exome sequencing of non-BRCA1/BRCA2 mutation carrier cases at high-risk for hereditary breast/ovarian cancer. Hum Mutat. 2021;42:290–9.
57. Helgadottir HT, Thutkawkorapin J, Lagerstedt-Robinson K, Lindblom A. Sequencing for germline mutations in Swedish breast cancer families reveals novel breast cancer risk genes. Sci Rep. 2021;11:14737.
58. Shagimardanova E, Shigapova LH, Enikeev RF, Druzhkov M, Nikitin A, Gusev O, et al. Landscape of germline mutations in hereditary breast and ovarian cancer (HBOC) patients in Russia revealed by target panel sequencing. Ann Oncol. 2018;29:viii76.
59. Su Y, Yao Q, Xu Y, Yu C, Zhang J, Wang Q, et al. Characteristics of Germline Non-BRCA mutation status of high-risk breast cancer patients in China and correlation with high-risk factors and multigene testing suggestions. Front Genet. 2021;12:674094.
60. Zhao Q, Yang J, Li L, Cao D, Yu M, Shen K. Germline and somatic mutations in homologous recombination genes among Chinese ovarian cancer patients detected using next-generation sequencing. J Gynecol Oncol. 2017;28:e39.
61. Gu Y, Jin Y, Li Y, Wang W, Shan Y, Xue C, et al. 853P The germline mutation landscape of DDR genes in Chinese patients with epithelial ovarian cancer. Ann Oncol. 2020;31:S635.
62. Guan X, Liao S, Zhang F, Zhu Q, Qiu H, Qin L, et al. Identifying the germline variation spectrum and predisposition genes in Chinese ovarian cancer using whole-exome sequencing. BMC Cancer. 2025;25:924.
63. Jin Y, Wang W, Wu M, Fan Y, Wang T, Huang C, et al. Clinicopathological features of Chinese ovarian cancer patients with double heterozygosity for cancer-predisposed genes. BMC Cancer. 2025;25:1391.
64. Li Y, Yang X, Cai H, Wang F. Homologous recombination deficiency among patients with germline or somatic non-BRCA1/2 homologous recombination repair gene variations. npj Precis Oncol. 2025;9:192.
65. Liu P, Lin C, Liu L, Lu Z, Tu Z, Liu H. RAD54B mutations enhance the sensitivity of ovarian cancer cells to poly(ADP-ribose) polymerase (PARP) inhibitors. J Biol Chem. 2022;298:102354.
66. Kottemann MC, Smogorzewska A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. Nature. 2013;493:356–63.
67. Mantere T, Tervasmäki A, Nurmi A, Rapakko K, Kauppila S, Tang J, et al. Case-control analysis of truncating mutations in DNA damage response genes connects TEX15 and FANCD2 with hereditary breast cancer susceptibility. Sci Rep. 2017;7:681.
68. Fu K, Li Q, Wang J, Zhang M, Yan X, Li K, et al. Characteristics of germline DNA damage response gene mutations in ovarian cancer in Southwest China. Sci Rep. 2024;14:6702.
69. Moes-Sosnowska J, Budziłowska A, Kupryjańczyk J. Mutation analysis of the FANCD2 gene in ovarian cancer patients from the Polish population. Nowotw J Oncol. 2015;65:7–13.
70. Adamson AW, Ding YC, Steele L, Leong LA, Morgan R, Wakabayashi MT, et al. Genomic analyses of germline and somatic variation in high-grade serous ovarian cancer. J Ovarian Res. 2023;16:141.
71. de Souza Timoteo AR, Gonçalves AÉMM, Sales LAP, Albuquerque BM, de Souza JES, de Moura PCP, et al. A portrait of germline mutation in Brazilian at-risk for hereditary breast cancer. Breast Cancer Res Treat. 2018;172:637–46.
72. Chevarin M, Alcantara D, Albuisson J, Collonge-Rame MA, Populaire C, Selmani Z, et al. The "extreme phenotype approach" applied to male breast cancer allows the identification of rare variants of ATR as potential breast cancer susceptibility alleles. Oncotarget. 2023;14:111–25.
73. Park, Odefrey DJ, Hammet FA, Giles F, Baglietto GG, Abcfs L, et al. FAN1 variants identified in multiple-case early-onset breast cancer families via exome sequencing: no evidence for association with risk for breast cancer. Breast Cancer Res Treat. 2011;130:1043–9.
74. Palovcak A, Yuan F, Verdun R, Luo L, Zhang Y. Fanconi anemia-associated protein 20 (FAAP20) plays an essential role in homology-directed repair of DNA double-strand breaks. Commun Biol. 2023;6:873.
75. Försti A, Ambrozkiewicz F, Marciniak M, Lubinski J, Hemminki K. Search for germline gene variants in colorectal cancer families presenting with multiple primary colorectal cancers. Int J Cancer. 2025;156:1393–403.
76. Fracassi G, Lorenzin F, Orlando F, Gioia U, D'Amato G, Casaramona AS, et al. CRISPR/Cas9 screens identify LIG1 as a sensitizer of PARP inhibitors in castration-resistant prostate cancer. J Clin Invest. 2024;135.
77. Rickman KA, Lach FP, Abhyankar A, Donovan FX, Sanborn EM, Kennedy JA, et al. Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia. Cell Rep. 2015;12:35–41.
78. Hira A, Yoshida K, Sato K, Okuno Y, Shiraishi Y, Chiba K, et al. Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia. Am J Hum Genet. 2015;96:1001–7.
79. Schultz-Rogers L, Lach FP, Rickman KA, Ferrer A, Mangaonkar AA, Schwab TL, et al. A homozygous missense variant in UBE2T is associated with a mild Fanconi anemia phenotype. Haematologica. 2021;106:1188–92.
80. Bogliolo M, Schuster B, Stoepker C, Derkunt B, Su Y, Raams A, et al. Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. Am J Hum Genet. 2013;92:800–6.
81. Levitus M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW, et al. The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. Nat Genet. 2005;37:934–5.
82. Torrezan GT, de Almeida F, Figueiredo MCP, Barros BDF, de Paula CAA, Valieris R, et al. Complex Landscape of Germline Variants in Brazilian Patients With Hereditary and Early Onset Breast Cancer. Front Genet. 2018;9:161.
83. Gómez-Flores-Ramos L, Barraza-Arellano AL, Mohar A, Trujillo-Martínez M, Grimaldo L, Ortiz-Lopez R, et al. Germline Variants in Cancer Genes from Young Breast Cancer Mexican Patients. Cancers (Basel). 2022;14.
84. Riedlova P, Janoutova J, Hermanova B. Frequency of mutations in BRCA genes and other candidate genes in high-risk probands or probands with breast or ovarian cancer in the Czech Republic. Mol Biol Rep. 2020;47:2763–9.
85. Velázquez C, KL, Esteban-Cardeñosa EM, Avila Cobos F, Lastra E, Abella LE, et al. Germline Genetic Findings Which May Impact Therapeutic Decisions in Families with a Presumed Predisposition for Hereditary Breast and Ovarian Cancer. Cancers. 2020;12.

86. Sopik V, Foulkes WD. Risky business: getting a grip on BRIP. J Med Genet. 2016;53:296–7.
87. Ramus SJ, Song H, Dicks E, Tyrer JP, Rosenthal AN, Intermaggio MP, et al. Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. J Natl Cancer Inst. 2015;107.
88. Weber-Lassalle N, Hauke J, Ramser J, Richters L, Groß E, Blümcke B, et al. BRIP1 loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer. Breast Cancer Res. 2018;20:7.
89. Peterlongo P, Figlioli G, Deans AJ, Couch FJ. Protein truncating variants in FANCM and risk for ER-negative/triple negative breast cancer. NPJ Breast Cancer. 2021;7:130.
90. Kiiski JI, Pelttari LM, Khan S, Freysteinsdottir ES, Reynisdottir I, Hart SN, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci USA. 2014;111:15172–7.
91. Kiiski JI, Tervasmäki A, Pelttari LM, Khan S, Mantere T, Pylkäs K, et al. FANCM mutation c.5791C>T is a risk factor for triple-negative breast cancer in the Finnish population. Breast Cancer Res Treat. 2017;166:217–26.
92. Nierenberg JL, Adamson AW, Hu D, Huntsman S, Patrick C, Li M, et al. Whole exome sequencing identifies FANCM as a susceptibility gene for estrogen-receptor-negative breast cancer in Hispanic/Latina women. Nat Commun. 2025;16:7816.
93. Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, Dao F, et al. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474:609–15.
94. Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, et al. Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. Cancer Cell. 2020;37:639–54.e6.
95. Stewart MD, Merino Vega D, Arend RC, Baden JF, Barbash O, Beaubier N, et al. Homologous Recombination Deficiency: Concepts, Definitions, and Assays. Oncologist. 2022;27:167–74.

ACKNOWLEDGEMENTS

We would like to thank Fiona Chan Pak Choon for her guidance with downstream exome analyses, Irving Simonin Wilmer for his guidance with statistical analyses in R, and Shelly Alkoby for her assistance with the curation of the framework gene panel. A previous version of this manuscript has been submitted as a preprint to *medRxiv* (https://doi.org/10.1101/2025.10.22.25338181).

AUTHOR CONTRIBUTIONS

JCV developed the computational genomic framework and bioinformatics pipelines and analyzed the data. TM, CR, and JLCF contributed to the pipelines' development. SG, CDRE, and PP contributed to the design of the statistical analyses. WDF, NH, CDRE, and PP designed the project. NH obtained access to controlled germline data from TCGA. PP provided curated promoter methylation data. WDF, NH, CDRE, PP, and BR provided supervision and scientific guidance. WDF, NH, and CDRE supervised all aspects of the project. JCV, WDF, NH, and CDRE wrote the manuscript. All authors reviewed the manuscript.

FUNDING

This work was supported in part by a grant provided to W.D.F. by the Canadian Institutes of Health Research (FDN-148390), as well as a fellowship for doctoral studies abroad from the National Council for Science and Technology (CONACYT) and an internal studentship for doctoral studies from the Research Institute of the McGill University Health Center (RI-MUHC) provided to J.C.V.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The TCGA consortium collected tumour and non-tumour biospecimens from participating human subjects under the authorization of the corresponding institutional review boards. Informed consent was obtained from all participants by the TCGA consortium. Further information is available at https://cancergenome.nih.gov/abouttcga/policies/informedconsent. Access to controlled germline sequencing data (blood-derived BAM files) from TCGA was obtained through the database of Genotypes and Phenotypes (dbGaP) and the GDC portal under an approved project (Project #34072; accession phs000178.v11.p8). Additional information regarding TCGA ethics and data use policies is available at: https://www.cancer.gov/ccg/research/genome-sequencing/tcga/history/ethics-policies. All methods were performed in accordance with the relevant guidelines and regulations. No identifiable human images are included in this study.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44276-026-00218-w.

Correspondence and requests for materials should be addressed to William D. Foulkes.