Structural Variants at the *BRCA1/2* Loci are a Common Source of Homologous Repair Deficiency in High-grade Serous Ovarian Carcinoma

Ailith Ewing[1], Alison Meynert[1], Michael Churchman[2], Graeme R. Grimes[1], Robert L. Hollis[2], C. Simon Herrington[2,3], Tzyvia Rye[2], Clare Bartos[2], Ian Croy[2], Michelle Ferguson[4,5], Mairi Lennie[5], Trevor McGoldrick[6,7], Neil McPhail[8], Nadeem Siddiqui[9], Suzanne Dowson[10], Rosalind Glasspool[11], Melanie Mackean[12], Fiona Nussey[12], Brian McDade[10], Darren Ennis[10,13], Lynn McMahon[14], Athena Matakidou[15], Brian Dougherty[16], Ruth March[17], J. Carl Barrett[16], Iain A. McNeish[10,11,13]; for the Scottish Genomes Partnership, Andrew V. Biankin[10,18,19], Patricia Roxburgh[10,11], Charlie Gourley[2], and Colin A. Semple[1]

ABSTRACT

Purpose: The abundance and effects of structural variation at BRCA1/2 in tumors are not well understood. In particular, the impact of these events on homologous recombination repair deficiency (HRD) has yet to be demonstrated.

Experimental Design: Exploiting a large collection of whole-genome sequencing data from high-grade serous ovarian carcinoma ($N = 205$) together with matched RNA sequencing for the majority of tumors ($N = 150$), we have comprehensively characterized mutation and expression at *BRCA1/2*.

Results: In addition to the known spectrum of short somatic mutations (SSM), we discovered that multi-megabase structural variants (SV) were a frequent, unappreciated source of *BRCA1/2* disruption in these tumors, and we found a genome-wide enrichment for large deletions at the *BRCA1/2* loci across the cohort. These SVs independently affected a substantial proportion of patients (16%) in addition to those affected by SSMs (24%), conferring HRD and impacting patient survival. We also detail compound deficiencies involving SSMs and SVs at both loci, demonstrating that the strongest risk of HRD emerges from combined SVs at both *BRCA1* and *BRCA2* in the absence of SSMs. Furthermore, these SVs are abundant and disruptive in other cancer types.

Conclusions: These results extend our understanding of the mutational landscape underlying HRD, increase the number of patients predicted to benefit from therapies exploiting HRD, and suggest there is currently untapped potential in SV detection for patient stratification.

Introduction

Homologous recombination deficiency (HRD) is identifiable in many cancers and is particularly prominent in high-grade serous ovarian cancer (HGSOC; ref. 1), affecting around 50% of tumors (2) and leaving detectable mutational spectra across the tumor genome (3). The mutational landscape of HGSOC is dominated by extensive genomic copy-number changes and structural rearrangement driven by chromosome instability and defective DNA repair, rather than the patterns of recurrent point mutation in tumor suppressor and oncogenes often observed in other solid tumors (4, 5).

Germline short mutations (GSM) disrupting the coding sequence of *BRCA1* and *BRCA2* are the most common types of HRD-associated defect, occurring in 8% and 6% of patients with HGSOC, respectively, while disruptive somatic short mutations (SSM) in these genes are present in an additional 4% and 3% of patients with HGSOC, respectively (6, 7). These GSMs and SSMs include single-nucleotide

[1]MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh, Scotland, United Kingdom. [2]Nicola Murray Centre for Ovarian Cancer Research, Cancer Research UK Edinburgh Centre, MRC IGMM, University of Edinburgh, Edinburgh, Scotland, United Kingdom. [3]Edinburgh Pathology, Cancer Research UK Edinburgh Centre, MRC IGMM, University of Edinburgh, Edinburgh, Scotland, United Kingdom. [4]Department of Oncology, Ninewells Hospital, NHS Tayside, Dundee, Scotland, United Kingdom. [5]Division of Molecular and Clinical Medicine, School of Medicine, University of Dundee, Dundee, Scotland, United Kingdom. [6]Department of Oncology, Aberdeen Royal Infirmary, Aberdeen, Scotland, United Kingdom. [7]Institute of Education for Medical and Dental Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, Scotland, United Kingdom. [8]Department of Oncology, Raigmore Hospital, NHS Highland, Inverness, Scotland, United Kingdom. [9]Department of Gynaecological Oncology, Glasgow Royal Infirmary, Glasgow, Scotland, United Kingdom. [10]Institute of Cancer Sciences, Wolfson Wohl Cancer Research Centre, University of Glasgow, Glasgow, Scotland, United Kingdom. [11]Beatson West of Scotland Cancer Centre and University of Glasgow, Glasgow, Scotland, United Kingdom. [12]Edinburgh Cancer Centre, Western General Hospital, NHS Lothian, Edinburgh, Scotland, United Kingdom. [13]Ovarian Cancer Action Research Centre, Department of Surgery and Cancer, Imperial College London, London, England, United Kingdom. [14]Precision Medicine Scotland (PMS-IC), Queen Elizabeth University Hospital, Glasgow, Scotland, United Kingdom. [15]Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, England, United Kingdom. [16]Translational Medicine, Oncology R&D, AstraZeneca, Waltham, Massachusetts. [17]Precision Medicine, Oncology R&D, AstraZeneca, Cambridge, England, United Kingdom. [18]West of Scotland Pancreatic Unit, Glasgow Royal Infirmary, Glasgow, Scotland, United Kingdom. [19]South Western Sydney Clinical School, Faculty of Medicine, University of NSW, Liverpool, New South Wales, Australia.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

P. Roxburgh, C. Gourley, and C.A. Semple contributed equally to this article.

Corresponding Author: Ailith Ewing, MRC Human Genetics Unit, University of Edinburgh, Edinburgh, Scotland EH4 2XU, United Kingdom. Phone: 4413-1651-8500; Fax: 4413-1651-8800; E-mail: ailith.ewing@igmm.ed.ac.uk

Clin Cancer Res 2021;27:3201–14

doi: 10.1158/1078-0432.CCR-20-4068

©2021 American Association for Cancer Research.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

Translational Relevance

We demonstrate that *BRCA1/2* losses by structural variation (predominantly large deletion) are novel and unappreciated events, leading to loss of gene expression in ovarian and other cancers. SVs contribute to homologous recombination repair deficiency and their detection could identify more patients suitable for PARP inhibition, a therapy with proven benefit.

variants (SNV), as well as short indels, with frameshifts being the predominant mechanism of inactivation. These *BRCA*-deficient tumors represent approximately 20% of patients with HGSOC. An additional 11% of tumors are thought to be *BRCA* deficient through epigenetic silencing of *BRCA1* (2, 8). Mutational or epigenetic inactivation of other genes involved in the homologous recombination (HR) pathway is also thought to confer HRD in a smaller proportion of patients with HGSOC (7, 9–12). Genome-wide patterns of SNVs, indels, and structural variation have been identified as strong predictors of *BRCA1/2* deficiency (3, 13). These mutational signatures of *BRCA1/2* deficiency are also found in additional tumors that lack short variants at *BRCA1/2*, suggesting that other unknown aberrations may also be involved in HRD (3). The demonstration of *BRCA1/2* loss and detection of HRD are crucial in the management of high-grade serous ovarian (14–19), breast (20, 21), pancreatic (22), and prostate (23) cancer to identify patients whose outcome is markedly improved by the administration of PARP inhibitors (14–16). PARP inhibitors selectively kill HRD cells because these cells are deficient in HR repair (HRR) and can neither resolve stalled replication forks nor accurately repair the increased number of double-strand breaks that result from the use of these agents (24).

The clinical importance of GSMs and SSMs at *BRCA1/2* is well established in cancer (17, 20, 25–28). In contrast, the abundance and effects of structural variants (SV) at *BRCA1/2* are not well understood, particularly for large SVs (>1 Mb) encompassing multi-megabase regions. Similarly, the compound effects of SVs and short mutations occurring simultaneously at *BRCA1* and *BRCA2* are poorly studied. Matched tumor-normal whole-genome sequencing (WGS) of freshly frozen tissue is accepted as the best resource to accurately detect SVs in tumors, but in the past such data have been scarce for HGSOC (29, 30). Here, we comprehensively characterize the mutational landscape of *BRCA1/2* in HGSOC using the largest collection to date of uniformly processed WGS data ($N = 205$), comprising two previously published cohorts (5, 6), as well as a large novel cohort described here for the first time. We document the prevalence of HRD across these three cohorts to reveal the complexity of the mutations associated with HRD, their impact on gene expression, and associations with clinical outcome.

Materials and Methods

WGS data from matched tumor and normal blood samples were uniformly remapped and analyzed to generate a range of somatic mutation calls (Supplementary Fig. S1) for three HGSOC cohorts: chemoresistant or relapsed tumors from the Australian Ovarian Cancer Study (AOCS; ref. 5; $N = 80$), pretreatment WGS primary tumors from The Cancer Genome Atlas (TCGA; ref. 6; $N = 44$), and the previously unpublished primary tumors from Scottish High-grade Serous Ovarian Cancer (SHGSOC) study ($N = 81$) of which 16 (20%) have had neoadjuvant chemotherapy. The combined cohort ($N = 205$) presented here was uniformly analyzed as follows.

Scottish sample collection and preparation for WGS

Scottish HGSOC samples were collected via local bioresource facilities at Edinburgh, Glasgow, Dundee, and Aberdeen and stored in liquid nitrogen until required. Patients with HGSOC were determined from pathology records and were included in the study where there was matched tumor and whole-blood samples. Tumor samples were divided into two for DNA and RNA extraction and slivers of tissue were taken, fixed in formalin, and embedded in paraffin wax (FFPE). Samples were only included if they were confirmed as HGSOC and there was greater than 40% tumor cellularity throughout the tumor, determined by using hematoxylin and eosin staining of the FFPE sections and pathology review. Somatic DNA was extracted from the tumor and germline DNA was extracted from whole blood. Quality control was then carried out on the resultant DNA to ensure sufficient purity and quality (Supplementary Materials and Methods). Only when all quality control requirements were satisfied, the DNA was sequenced at the Glasgow Precision Oncology Laboratories.

This study was carried out in accordance with the principles of the Declaration of Helsinki. Ethical approval for the use of human tissue specimens for research was obtained from the Lothian NHS Research Scotland Human Annotated Bioresource (ethics committee reference, 15/ES/0094-SR494). Correlation of molecular data to clinical outcome and clinicopathologic variables in ovarian cancer was approved by South East Scotland Research Ethics Committee 2 (reference, 2007/W/ON/29). All relevant ethical regulations were complied with, including the need for written informed consent where required.

Sequence acquisition and availability

WGS and RNA sequencing (RNA-seq) reads were downloaded in compressed FASTQ format from the sequencing facility (SHGSOC) or in aligned BAM format (including unaligned reads) from the European Genome/Phenome Archive (AOCS) and the Bionimbus Protected Data Cloud (TCGA_US_OV). The reads obtained in BAM format were query sorted and converted to FASTQ. All whole-genome sequence and RNA-seq data for the SHGSOC cohort will be made available on publication via European Genome-Phenome Archive (EGA).

Primary processing of WGS

Reads were aligned to the GRCh38 reference genome and somatic and germline variant calling was run using a bcbio (31) 1.0.7 pipeline (see Supplementary Data for full pipeline configuration, program, resource versions, and references). Germline SNPs and indels were called with GATK 4.0.0.0 HaplotypeCaller. Somatic SNVs and indels were called as a majority vote between Mutect2, Strelka2, and VarDict. Small variants were annotated with Ensembl Variant Effect Predictor v91 and filtered for oxidation artefacts by GATK 4.0.0.0 FilterByOrientationBias. Somatic SVs were called with Manta 1.2.1 and somatic copy-number variants with CNVkit 0.9.2a0. LOH and somatic copy-number variants were also identified with CLImAT. Sample quality control was performed with Qsignature 0.1 to identify sample mix-ups and with VerifyBamId 1.1.3 to identify sample contamination. Tumor cellularity was estimated using both CLImAT's estimates and p53 variant allele frequency. These measures were compared with the qPure estimates for the AOCS cohort (5) and histopathologic estimates for the SHGSOC cohort with very good concordance (Supplementary Fig. S2). The CLImAT estimates were used as the final estimates of cellularity.

Filtering of small variants (SNVs and indels) at BRCA1/2

Germline short variants at *BRCA1/2* were filtered to include only damaging pathogenic variants for the purposes of establishing *BRCA1/2* mutational status. Included variants were all of moderate

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

or high impact according to variant effect predictor (VEP). Variants with a pathogenic or risk factor annotation according to ClinVar (32) were included ($n = 145$). Remaining variants with a ClinVar benign or likely benign status were excluded ($n = 1{,}147$). Remaining frameshift or nonsense (stop gained) or splice donor/acceptor variants were included ($n = 125$). Remaining missense variants with damaging SIFT and PolyPhen predictions were included ($n = 36$). Remaining missense variants called as damaging by only one of SIFT and PolyPhen were considered borderline and were excluded if their CADD score < 20 (33). Missense or in-frame variants with no Clinvar, SIFT, or PolyPhen evidence were excluded.

Somatic short variants at *BRCA1/2* were also filtered for pathogenicity to include variants that: were annotated by VEP as being of high or moderate impact, were pathogenic according to at least one of SIFT or PolyPhen, and had a high CADD score. In addition, we excluded somatic variants with an allele frequency less than 0.4.

Curating a high-confidence list of SVs at BRCA1/2

We identified SVs in patients with HGSOC using Illumina paired and split read–based SV detection tool, Manta (34). However, we observed that Manta was failing to detect a large number of very large deletions (>1 Mb) that had been identified using depth of coverage–based approaches in the Pan-Cancer Analysis of Whole Genomes (PCAWG; ref. 35). Therefore, we chose to supplement these calls with copy-number variants greater than 1 Mb in size that were called by one caller (CNVkit) using evidence from read depth and were also confirmed by an allele-specific copy-number caller (CLImAT), which provides an additional layer of evidence in addition to read depth as it also incorporates the shift in allele frequency of heterozygous SNPs within the potential copy-number variant into its variant calling algorithm, which also accounts for aneuploidy and sample cellularity. Deletions were assumed to be heterozygous if the copy number as estimated by both CNVkit and CLImAT was 1, although this may be a conservative estimate of allelic loss in the presence of subclones. We visually inspected all the identified SVs in the Integrative Genomics Viewer (v2.4.10). The magnitude of coverage and log fold change were inspected to confirm either duplication or deletion. For SVs in the range 300 bp–30 kb, the paired-end sequencing reads were manually reviewed, including looking at split reads, paired-end insert size, read coverage, and pair-orientation. We compared our filtered set of large copy-number variations (CNV) in the samples included in PCAWG with PCAWG's copy-number calls and found that 40 of 41 of our variants in these samples were also identified by PCAWG.

Enrichment of structural variation at *BRCA1/2*

Permutation analyses were carried out using the R package RegioneR (36) to investigate whether large deletions overlap more often with *BRCA1* and *BRCA2* than they do elsewhere in the genome. We carried out 100,000 permutations to simulate the null hypothesis throughout the genome for each gene and judged significance at alpha = 0.05 (Supplementary Fig. S3).

We investigated whether large deletions, large duplications, and inversions are enriched at *BRCA1* and *BRCA2* within their respective chromosomes using 100,000 circularized permutations to generate the null distribution of overlaps in each case (Supplementary Fig. S4). The observed number of overlaps with *BRCA1/2* was well within the range of this null distribution and, therefore, showed no evidence of within-chromosome enrichment, which is perhaps unsurprising given the large sizes of these events relative to the length of their chromosomes.

Implementation of HRDetect

To predict the level of HR deficiency in each tumor sample, we implemented the HRDetect algorithm as published by Davies and colleagues (3). We based our implementation on a Snakemake pipeline made publicly available by Zhao and colleagues (37), with some modifications to ensure accurate recapitulation of the original method (Supplementary Materials and Methods). As some of the AOCS cohorts' patients included here were also used in the validation of HRDetect in the original publication, we were able to compare our implementation for the same patients with that of the authors. Our implementation of HRDetect was very highly correlated with the original HRDetect implementation on the same samples (Spearman rho = 0.92).

We used the weights for the independent variables that were defined by the original model rather than retraining the model on our data as the original weights trained on a breast cancer dataset have been shown to perform well on ovarian cancer datasets and our total sample size was substantially smaller than that used to train the model originally. As input, we used the somatic SNVs and indel calls identified by the ensemble calling approach described above, with the SV calls made by Manta and the copy-number segments defined by CNVkit.

Scottish RNA sample preparation and sequencing

HGSOC samples were collected and underwent quality control as described for the DNA samples used for WGS. Somatic RNA was extracted from the resulting RNA sample, underwent quality control, and was quantified. RNA-seq was carried out by the Edinburgh Clinical Research Facility (Edinburgh, Scotland, United Kingdom) on an Illumina NExtSeq500 (further details in Supplementary Materials and Methods).

Primary processing of RNA-seq

RNA-seq data were analyzed using the Illumina RNA-seq best practice template. Briefly, reads were aligned to the GRCh38 reference genome and quality control was carried out. Salmon quant (38) was used to quantify the expression of transcripts against the GRCh38 RefSeq transcript database indexed using the salmon index (k-mers of length 31). Transcript-level abundance estimates were imported into R and summarized for further gene-level analyses. For differential expression analyses, raw expression counts were used by the DESeq2 package (39). For visualization of gene expression, counts were normalized using the variance stabilizing transformation. Previously published RNA-seq data available for the AOCS (ref. 5; $N = 80$) and TCGA (ref. 6; $N = 30$) cohorts, together with novel RNA-seq data for the SHGSOC ($N = 40$) cohort, generated for this study as detailed above, were processed in this way from FASTQ.

Curation and acquisition of the patient's clinical information

SHGSOC

Clinical data for the SHGSOC cohort were retrieved from the Edinburgh Ovarian Cancer Database (40), the CRUK Clinical Trials Unit Glasgow, and available electronic health records (ethics reference, 15/ES/0094-SR751).

AOCS and TCGA

The clinical information, including survival endpoints, age, and stage at diagnosis, is available for these patients as part of the PCAWG project (35).

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

Statistical analyses

All downstream statistical analyses were carried out in R (v3.6.0) using Jupyter notebook (v4.3.1).

Differential expression analyses of BRCA1/2 in tumors with and without deletions at BRCA1/2

To compare the gene expression levels at *BRCA1/2* between tumors with and without *BRCA1/2* deletions, we used the package DESeq2 to test for differential expression between the raw gene expression counts at each gene between samples with and without a deletion at that gene. At *BRCA1*, samples that also had a short variant had significantly lower expression than those that had a deletion alone. As a result, we only considered the samples with a deletion in the absence of a short variant. At *BRCA2*, this was not the case and the samples with short variants in addition to a deletion had comparable levels of expression with those with deletions alone, and were therefore included in the analysis. Cohort and tumor sample cellularity were included as covariates in the model formula.

Univariable analyses of genomic features and risk of HR deficiency

The risk of HR deficiency in tumors with *BRCA1/2* mutations, grouped by type, relative to those tumors without *BRCA1/2* mutations, was determined by using Fisher exact tests. The effect of mutations at *BRCA1* and *BRCA2* was determined together and, where sample size permitted, separately for mutational categories including: GSMs only, SSMs only, single deletion which overlaps at least one exon in the absence of a short variant or other SVs, deletion of at least one exon at both *BRCA1* and *BRCA2* in the absence of a short variant, inversion spanning *BRCA1* in the absence of short variants, or deletion and duplication spanning *BRCA2* in the absence of short variants or deletion. We also considered the impact of the presence of a short variant accompanied by a deletion at either of the genes. These categories are further described together with their labels and color coding in the Supplementary Data. The RR conferred by each mutational category was calculated in comparison with the group of patients without *BRCA1/2* mutations or SVs. Samples where *BRCA1/2* promoter methylation had been detected were excluded, except for where the effect of BRCA1 promoter methylation was itself being examined. All samples with *BRCA1* promoter methylation are predicted to be HR deficient so pseudocounts of 1 are used to estimate the effect size which is, therefore, a likely underestimate. *P* values were adjusted for the impact of multiple testing using Benjamini–Hochberg correction and were considered together with the effect sizes in the reporting of results.

Multivariable elastic net regularized regression model

Given the relative sparsity of the data and the correlation between features, we used a multivariable elastic net regularized regression model for the binary outcome of HRD defined by a probability of HRD greater than 0.7 from HRDetect. The data were partitioned into train and test sets (80:20) and the tuning parameters were optimized, to maximize the AUC, using 10-fold cross-validation of the training set. The model was then fitted to the training set and the model performance was assessed using the test set. The input variables available for selection were: *BRCA1* germline short variant status, *BRCA1* somatic short variant status, the presence of a large somatic deletion at *BRCA1*, the presence of a large somatic deletion at *BRCA1* and a *BRCA1* short variant, all the corresponding variables for *BRCA2*, the presence of an inversion at *BRCA1*, the presence of a duplication at *BRCA2*, the presence of a large somatic deletion at *BRCA1* and at *BRCA2* (double deletion), *BRCA1* promoter hypermethylation, whole-genome doubling, genome-wide load of SNVs, large CNVs, and SVs, in addition to cohort and tumor cellularity. The data were partitioned and the model was optimized and fitted to 100 train-test splits of the data to assess the robustness of the feature selection.

Survival-time analyses of the impact of HRD on overall survival

Follow-up information, including overall survival time, was available for 190 of 205 patients, of which 144 were deceased by the time of last follow-up. The association between genome-wide patterns of HRD and progression-free survival (PFS) was also assessed. Progression-free interval time was available for 151 of the patients from the AOCS and SHGSOC cohorts, of which 129 relapsed by the time of last follow-up. The effect of the HRDetect score, as a measure of the probability of HRD in the tumor, on the length of time that patients survived after diagnosis (overall survival time) and the time between diagnosis and first radiologically defined progression (PFS time) was assessed using Cox proportional hazards models stratified by cohort. Multivariable models were also fitted adjusting for age and stage at diagnosis and the Schoenfeld residuals were examined. In all cases, hazard ratios reported corresponded to a one SD increase in HRDetect score. Survival probability through time was compared between HR-deficient (HRDetect score > 0.7) and HR-proficient (HRDetect score ≤ 0.7) patients in Kaplan–Meier plots. This was repeated excluding the patients with *BRCA1/2* short variants to assess the impact of HRD driven by other events on survival.

BRCA1/2 SVs in other cancer types and their impact on expression

We used the consensus SV and CNV calls generated and included in the 2017-01-19 release of PCAWG to investigate the abundance of SVs at *BRCA1/2* in other cancer types. Samples ($N =$ 2,567) were from all cancer subtypes included in PCAWG that are recommended for use in subtype-specific analyses. Disruptive SVs were identified by intersecting the variant calls with the exons of *BRCA1/2*. Abundance of SVs, deletions, and duplications were tabulated separately for each of the cancer types. Cancer subtypes were ordered from left to right by overall SV burden as determined by PCAWG.

Using gene expression quantifications from matched RNA-seq for $N = 735$ of these samples, we compared *BRCA1* expression in samples with and without *BRCA1* deletions from each cancer type separately in differential expression analyses carried out using DESeq2 (39). Cancer types were included if there were greater than 15 samples in total and if at least three of these samples harbored deletions. This effect was also measured across all samples pan-cancer adjusting for differences in expression between cancer types in the model. These analyses were repeated for *BRCA2* expression. Effects on expression and the data behind them were displayed after variance stabilizing transformation of the expression counts.

Data availability

Previously published WGS and RNA-seq data that were reanalyzed here are available via EGA at accession code EGAS00001001692 (ICGC PCAWG). WGS, RNA-seq, and clinical data from the Scottish cohort (SHGSOC) will be made available via EGA at accession code EGAS00001004410. Other supporting data have been provided in the Supplementary Tables.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

Code availability

All codes will be made available at: https://github.com/ailithewing/Structural_variants_BRCA1_2_HRD_inHGSOC.

Results

Large SVs are a frequent source of *BRCA1/2* disruption in HGSOC

A variety of SVs were detected at the *BRCA1/2* loci, but large multi-megabase deletions spanning the entirety of *BRCA1* or *BRCA2* dominated. In all three cohorts, some deletions encompassed more than 10% of chromosomes 17 or 13, although the majority were more focal (median *BRCA1* deletion = 4.9 Mb and median *BRCA2* deletion = 6.2 Mb; **Fig. 1**). Heterozygous deletions occurred at similar rates at *BRCA1* (16%) and *BRCA2* (14%) overall, and at comparable rates between the similarly sized AOCS and SHGSOC cohorts (Supplementary Table S1). In six of 205 (3%) samples in the combined cohort, we observed large deletions at both *BRCA1* and *BRCA2* in the absence of short mutations at both genes. Permutation analyses revealed that *BRCA1* and *BRCA2* were deleted more than expected given the observed distribution of large deletions throughout the HGSOC genome (Supplementary Fig. S3). This is consistent with selection for these events as has been reported for SNVs at *BRCA1/2* (41), but we cannot exclude the role of mutational bias in producing these enrichments of deletions. Inversions occurred less often than deletions, but did occur in isolation in 6% of samples, and within groups of large

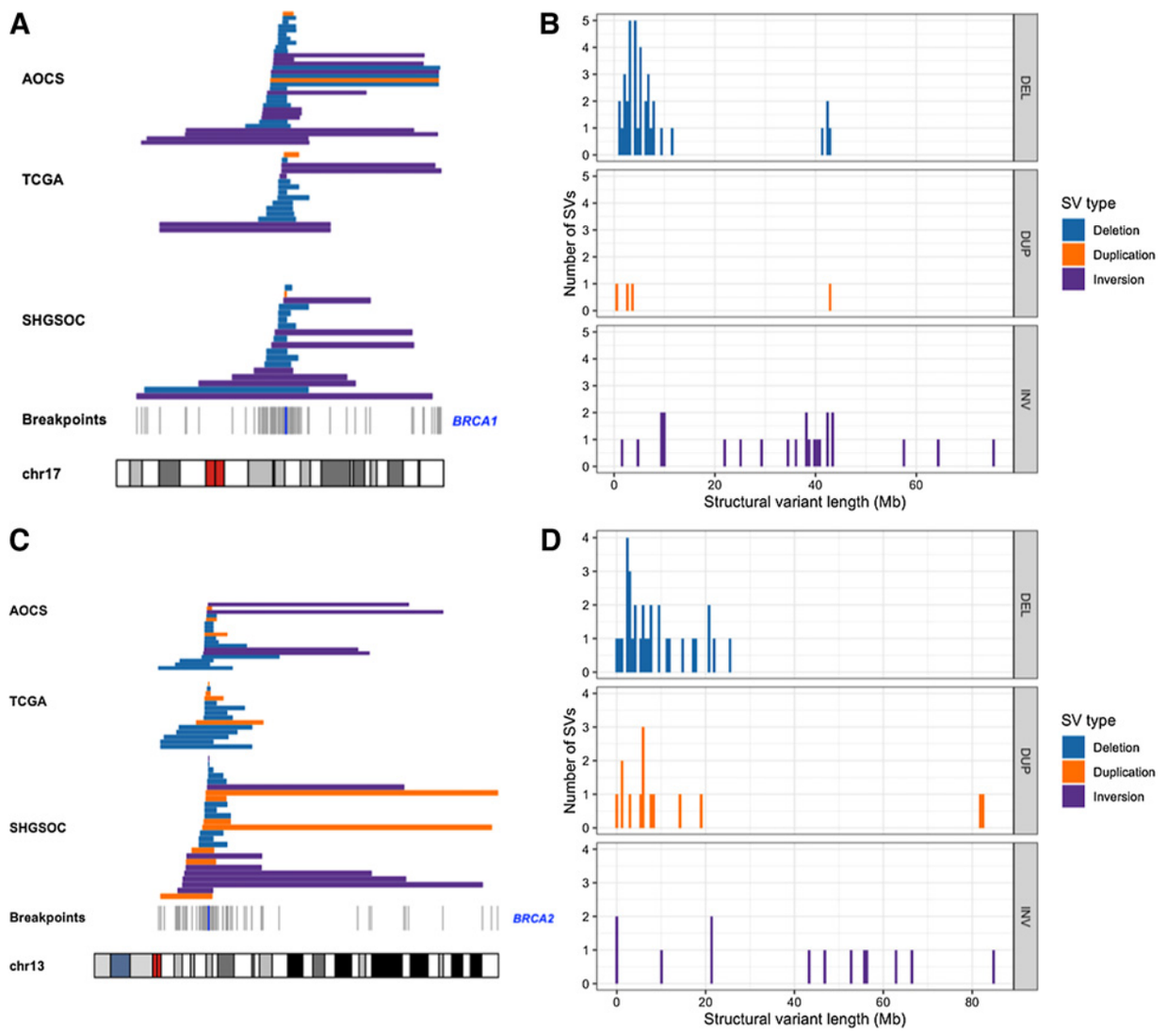

Figure 1.

Abundance, location, and size of SVs overlapping *BRCA1/2* in three HGSOC cohorts. **A,** Alignment of SVs overlapping *BRCA1* across the AOCS, TCGA, and SHGSOC cohorts with breakpoints marked in gray according to their position on chromosome 17. Location of *BRCA1* marked by a blue line with deletions (blue), duplications (orange), and inversions (purple). **B,** The distribution of sizes of SVs (Mb), overlapping *BRCA1* across all cohorts. **C,** Alignment of SVs overlapping *BRCA2* across the three cohorts with breakpoints marked in gray according to their position on chromosome 13. Location of *BRCA2* marked by a blue line. **D,** The distribution of sizes of SVs (Mb) overlapping *BRCA2* across all cohorts with deletions (blue), duplications (orange), and inversions (purple).

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

overlapping inversions in 5% of samples. In addition, we observed large duplications that spanned the entire length of either *BRCA1* or *BRCA2* in all cohorts (2% and 7%, respectively; Supplementary Table S2). LOH was near ubiquitous at *BRCA1* (202/205; 99%), of which 166 events were copy-number neutral, and was present at *BRCA2* in more than half of samples (118/205; 58%), of which 91 events were copy-number neutral. We observed similar genome-wide SV mutational spectra in each cohort despite the clinical differences among them: in particular AOCS represents chemoresistant/relapsed cases, while SHGSOC is composed mainly of samples taken before treatment. This suggests that large SVs predicted to impair *BRCA1/2* function are a general feature of HGSOC evolution.

Large deletions spanning *BRCA1/2* are associated with lower gene expression

We found that patients with a large deletion spanning *BRCA1*, in the absence of a GSM or SSM, had lower *BRCA1* expression than those patients with no *BRCA1* GSM, SSM, or SVs ($\log_2$ fold change of no GSM/SSM/SVs vs. a deletion only = 0.45; $P = 0.0093$; **Fig. 2**). We also found that tumors with a large deletion spanning *BRCA2* had lower *BRCA2* expression than those without GSM/SSM/SVs at *BRCA2* ($\log_2$ fold change in expression between no GSM/SSM/SVs vs. a deletion = 0.43; $P = 0.037$). In the analysis of *BRCA2* expression, we included tumors with large deletions that also had GSM or SSM in the remaining copy, having shown that in our data, *BRCA2* expression is not significantly different in these tumors (Supplementary Fig. S5). In spite of the unavoidable heterogeneity in tumor expression data among samples, although variation in tumor purity was accounted for, the trends observed here are consistent with large *BRCA1/2* deletions, reducing *BRCA1/2* expression, although indirect mechanisms cannot be excluded.

Large deletions spanning *BRCA1/2* contribute to HRD independently of pathogenic SNVs and indels

We examined the functional impact of all *BRCA1/2* mutations detected across all cohorts using an established method, HRDetect (3), which predicts HRD based upon genome-wide mutational spectra, and, therefore, provides a functional readout for the HRR pathway in tumors (**Fig. 3**).

As expected, tumors with GSMs or SSMs in *BRCA1/2* are more likely to have HR-deficient tumors (HRDetect scores > 0.7) than those tumors without GSMs, SSMs, or SVs at these genes [GSM OR, 6.9; 95% confidence interval (CI), 1.8–33; $P = 2.3 \times 10^{-3}$; $P_{adj} = 3 \times 10^{-2}$ and SSM OR, 25; 95% CI, 3.2–1,121; $P = 1.7 \times 10^{-4}$; $P_{adj} = 2.3 \times 10^{-3}$; **Fig. 3**). Four samples with GSMs at *BRCA2* demonstrated low HRDetect scores and accordingly showed no evidence for subsequent loss of the wild-type allele in the tumor which suggests that certain GSMs that are predicted to be disruptive are insufficient to generate HRD (Supplementary Table 3a).

The majority of samples (65%) with *BRCA1/2* deletions were HRD (HRDetect score > 0.7), although some deletions occurred in tumors that also harbored short mutations (19/49). Because many of the samples with *BRCA1/2* deletions lacked short mutations (30/49; 61%), analysis of the effects of deletions independent of short variants was undertaken. The compound effect of deletions at both the *BRCA1* and *BRCA2* loci in the absence of short variants was particularly pronounced, demonstrating a significantly increased risk of HRD (OR, 19; CI, 2.4–896; $P = 1.3 \times 10^{-3}$; $P_{adj} = 1.7 \times 10^{-2}$) and consistently high HRDetect scores. Furthermore, compound deletions at both loci generated an OR that was comparable with other classes of HRD mutations known to have clinical importance, including the well-studied disruptive *BRCA1/2* short variants (**Fig. 3C** and **D**). Single deletions at either *BRCA1* or *BRCA2* did not consistently confer an increased risk of HRD in the absence of a *BRCA1/2* short variant (**Fig. 3**), although the analysis may be underpowered to detect small effects given the current sample size. The HRDetect scores for samples with these single deletions formed a bimodal distribution for which we have been unable to find a defining characteristic for the difference, such as the length of the deletion, resultant level of gene expression, or a background of whole-genome doubling. The estimated effect that we observe of a single deletion at *BRCA2* merits further investigation, although it did not achieve statistical significance in our data (OR, 2; 95% CI, 0.37–10.3). Previous studies have identified rearrangement signatures associated with HR deficiency (42). We observed an elevation in these signatures, particularly in the case of rearrangement signature 5, in samples with deletion at *BRCA1/2*, which is broadly consistent with the levels of elevation observed in the presence of short mutations at *BRCA1/2* (Supplementary Fig. S6). However, HRDetect incorporates additional genome-wide sources of information, including the more powerful presence of indel microhomology leading to more accurate HRD prediction.

Beyond deletions, the functional impact of other classes of SV, such as inversions or duplications, is less well studied. Half of the samples with only *BRCA1* inversions bear an HR-deficient signature, which suggests that although in isolation their presence is not associated with HR deficiency, these samples can show evidence of HRD perhaps via another unappreciated route (**Fig. 3**). In contrast, only one of the samples with only *BRCA2* duplications was HR deficient, which suggests potential for enrichment in HR-proficient samples, but this would need to be further explored in greater sample sizes.

Deletions are a frequent source of *BRCA1/2* inactivation in repair deficiency

Samples across the combined cohort never had more than one short mutation across *BRCA1* and *BRCA2*. This suggests that SSMs are not a mechanism for biallelic inactivation of a gene affected by a GSM. This is consistent with reports from a previous study of HGSOC (ref. 41; Supplementary Table S3A and S3B). In contrast, of the HGSOC tumors with a GSM at *BRCA1/2* predicted to cause HRD, we found that 11 of 32 (34%) showed evidence for an SV at the same gene which may contribute to HRD if the SV occurs on the other allele (Supplementary Tables S2 and S3). We found that most of these somatic events (8/11; 73%) were large deletions, while two tumors possessed more than one SV spanning the same gene as the GSM and one more showed evidence of somatic duplication. The importance of "second hit" (43) mutations in tumors is well established (44), but these data suggest that multi-megabase deletions have an underappreciated role in this phenomenon in HGSOC. Across the three cohorts, 24% (50/205) of patients had a disruptive short variant at either *BRCA1* or *BRCA2*, and 30% (15/50) of these patients also carried a *BRCA1/2* deletion at the same locus. Also, it appears that SVs, including deletions, can occur at both *BRCA1* and *BRCA2* in the same sample. Large somatic deletions occurred at both *BRCA1* and *BRCA2* in 13 samples and in seven of these samples, there was no short variant at either gene, although one sample had a hypermethylated *BRCA1* promoter. These data suggest that large deletions and other SVs disproportionately contribute to biallelic inactivation in HGSOC, driven by the unusually high rates of structural variation seen in this cancer (45).

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

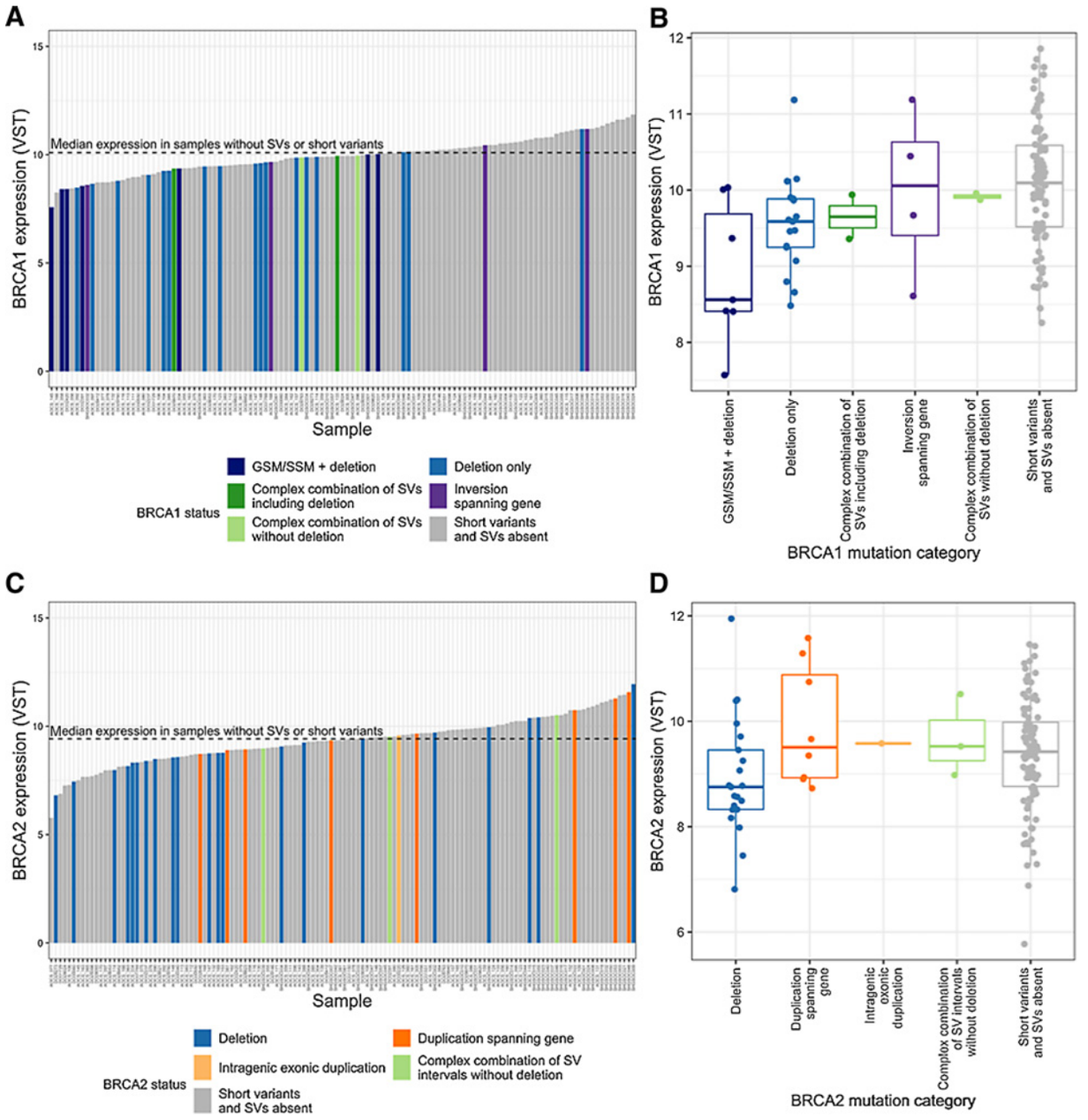

Figure 2.

Structural variation and expression of *BRCA1/2* in the combined HGSOC cohort. **A** and **C,** Expression of *BRCA1/2* (variance stabilizing transformed RNA-seq counts) across samples, ordered from lowest to highest expression. Median *BRCA1/2* expression is indicated by a black dashed line. Sample bars are colored by *BRCA1/2* mutational category. **B** and **D,** Box plot of *BRCA1/2* expression for each category of *BRCA1/2* mutation. The *BRCA1* deletion category is split into those samples with SNVs and deletions and those with only deletions as their expression is significantly different (Supplementary Fig. S5A). This was not the case for *BRCA2*, so all samples with deletions are considered together to maximize the available sample size (Supplementary Fig. S5B).

Integrative modeling reveals complex mechanisms underlying repair deficiency

We comprehensively modeled the effects of a range of genomic alterations at the *BRCA1/2* loci on HRD, to investigate the relative importance of these features in explaining the patterns of HRD observed. Given the relative sparsity of the data and the correlation between features, we used a multivariable elastic net regularized regression model.

In addition to the previously reported impact of short variants at *BRCA1/2* and *BRCA1* promoter hypermethylation on HRD, large deletions at *BRCA2* conferred an increased risk of HRD (**Fig. 4**). Furthermore, samples with double deletions, where deletions were found at both *BRCA1* and *BRCA2*, were more likely to be HR deficient. Importantly, the influence of these double deletions on HRD exceeded that of genome-wide large CNV loads and genome-wide SV loads. Also, large inversions

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

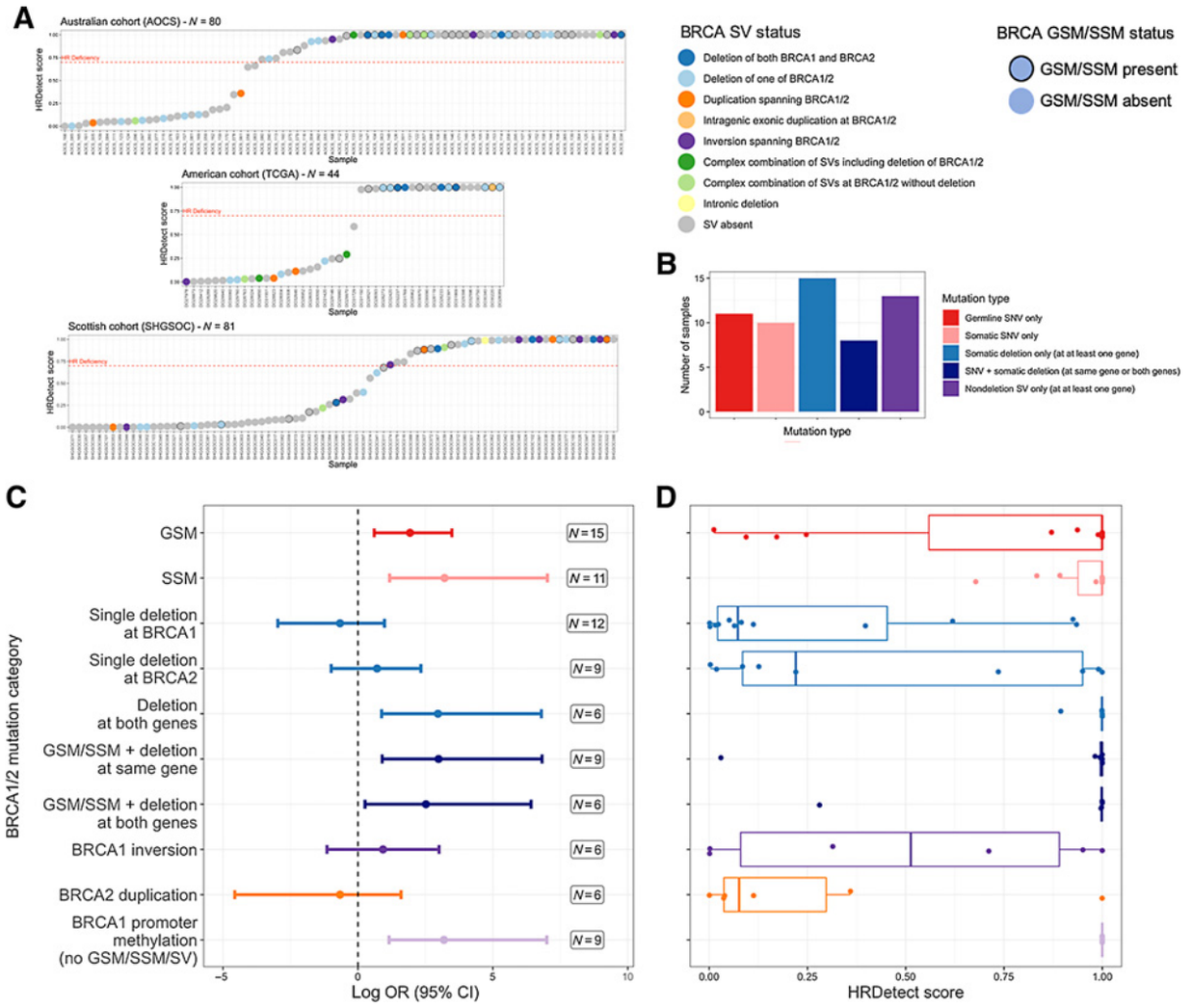

Figure 3.

BRCA1/2 mutation classes and repair deficiency in three HGSOC cohorts. **A,** Predictions of HRD for three large cohorts of HGSOC colored by *BRCA1/2* SV status and outlined by BRCA1/2 mutation status. Categories of mutation include GSMs and SSMs. Categories of structural variation include deletions, duplications, inversions, and complex overlapping combinations thereof as formally described in the Materials and Methods and absence of SVs at *BRCA1/2*. The HRDetect scores range from 0, least likely to be HR deficient, to 1, most likely to be HR deficient. The red dashed line represents the threshold of 0.7 representing HRD (3). **B,** The number of HRD tumors with different categories of *BRCA1/2* short variants, deletions, or nondeleting SVs. **C** and **D,** The increase in log OR of HRD (HRDetect score > 0.7) associated with different categories of mutation and structural variation at *BRCA1/2* in comparison with the frequency of the reference category, where samples lack evidence of *BRCA1/2* inactivation (GSV, SSV, and SV). All categories, apart from the BRCA1 promoter methylation category itself, also exclude tumors with *BRCA1* promoter methylation where this is known (TCGA and AOCS). ORs are defined using Fisher exact tests for enrichment. Error bars represent 95% CIs. Mutually exclusive categories of mutation examined include GSV only, SSV only, the presence of a deletion at one or both genes without a GSV or SSV, the presence of a short variant together with deletion of one or both genes, nondeleting SVs in samples without short variants or deletions, samples with *BRCA1* promoter methylation, and no mutational *BRCA1/2* deficiencies.

at *BRCA1* were independently associated with an increased risk of HRD. The functional impact of these events on the gene is currently unknown, but this suggests that these events may either be markers for processes that impact the gene's function or may even directly impact the function of the gene themselves. The model's ability to predict HRD was good, with a mean ROC curve AUC of 0.75, which although promising, suggests that there are additional unknown sources of HRD.

We can explain the observed pattern of HRD by the presence of mutational or epigenetic defects at the *BRCA1/2* genes in 81 of 106 samples with predicted HRD (72 GSM/SSM/SVs at *BRCA1/2*, nine with *BRCA1* promoter methylation), but a further 25 samples with HRD remain unexplained. On further examination, we found that all of these samples harbored damaging GSMs and/or SSMs at other HR genes (defined by Kyoto Encyclopedia of Genes and Genomes pathway annotation; Supplementary Tables S5–S7), motivating analysis of the potential roles of mutations at loci other than *BRCA1/2* and their inclusion in an expanded model. However, we found no convincing evidence for a strong influence of mutations at other HR genes or the combined expression of genes dysregulated in the presence of HRD (Supplementary Fig. S7).

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

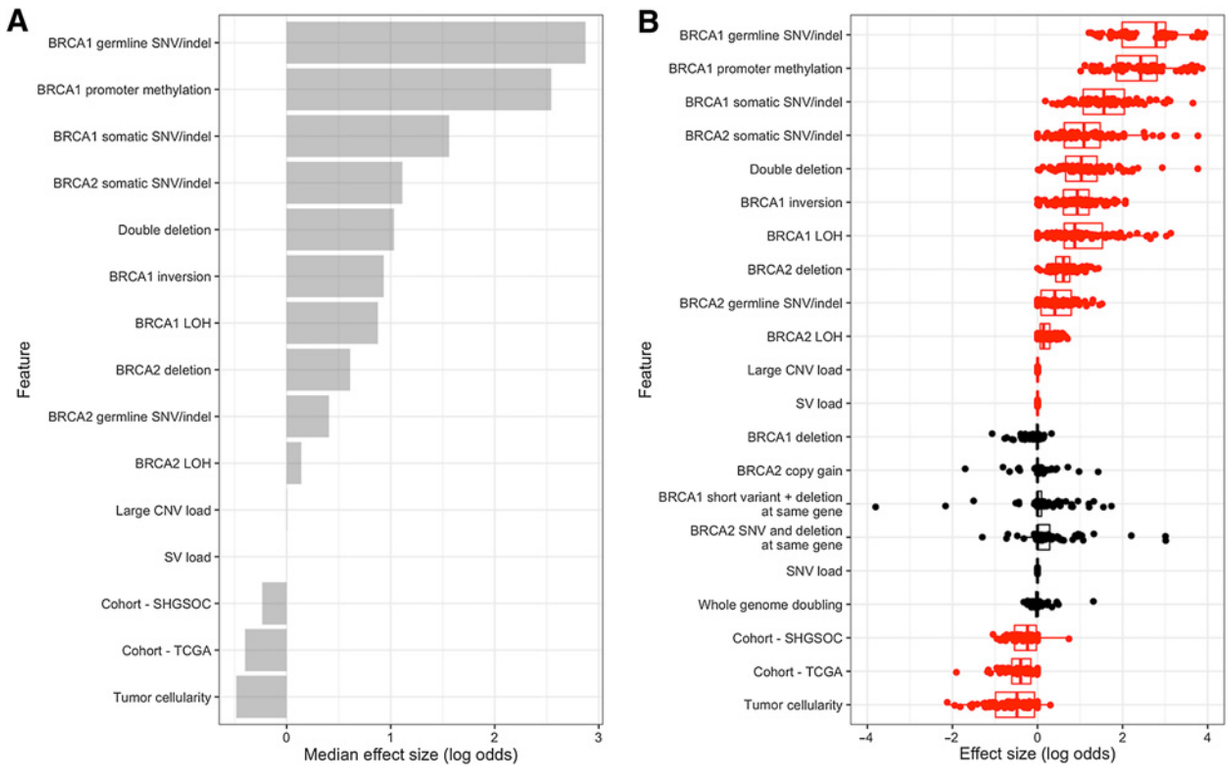

Figure 4.

Integrative modeling of repair deficiency in HGSOC. **A,** Median effect sizes of genomic features selected to predict HRD, using elastic net regularized regression on 100 training/test set splits. Model performance was measured for each split and average AUC = 0.75. Binary mutational status variables (e.g., presence/absence of *BRCA1* somatic SNV) were included as factors and continuous variables were standardized to allow comparisons between variables. **B,** Distributions of effect size for each variable on HRD (log odds) in each training/test set split. Variables in red were selected for inclusion by the model in more than half of the training sets.

HRD is associated with longer survival in the absence of disruptive short variants at *BRCA1/2*

In HGSOC, while HRD as a result of GSMs and SSMs at *BRCA1/2* is associated with response to platinum and PARP inhibition and improved survival, the relationship between HRD resulting from disruption of *BRCA1/2* via other mechanisms is less clear (1, 6, 8, 11, 46–49).

A higher probability of HRD was significantly associated with longer overall survival in our combined cohort [Hazard Ratio, 0.64; 95% CI, 0.53–0.76 (per 1 SD increase); $P = 8.4 \times 10^{-7}$] and this effect was only slightly attenuated by adjustment for patient age and tumor stage at diagnosis (**Fig. 5A**). Notably, this effect persisted [Hazard Ratio, 0.67; 95% CI, 0.56–0.84 (per 1 SD increase); $P = 3.6 \times 10^{-4}$; **Fig. 5B**] when we excluded the patients with *BRCA1/2* GSM/SSM, who are already known to have longer survival. We saw a similar effect when we examined the effect of HRD on PFS [Hazard Ratio, 0.77; 95% CI, 0.64–0.93 (per 1 SD increase); $P = 0.007$], which also was robust to the exclusion of patients with *BRCA1/2* GSM/SSM [Hazard Ratio, 0.79; 95% CI, 0.64–0.98 (per 1 SD increase); $P = 0.03$]. Correspondingly, in all of these instances, patients with HRD tumors (HRDetect > 0.7) had longer survival times than patients with non-HRD tumors. The effects observed are consistent with the presence of HRD through mechanisms other than *BRCA1/2* GSMs and SSMs affecting overall survival.

Large deletions at BRCA1/2 are abundant in other cancer types and disrupt expression

Cancers with other sites of origin vary in their level of structural variation. It is possible that large deletions may compromise *BRCA1/2* function in these cancers also. The PCAWG consortium recently generated conservative consensus SV calls based upon WGS data across many tumor types, but did not report specifically on SV patterns at *BRCA1/2* (50). These data illustrate that large deletions (**Fig. 6A**; and other SVs; Supplementary Fig. S8) at *BRCA1/2* are common across a variety of non-HGSOC cancer types, including in cancers also known to show evidence of HRD, such as breast and prostate cancer. In addition, two rare tumor types showed notably higher frequencies of *BRCA1/2* deletions: chromophobe renal cell carcinoma (CHRCC) and leiomyosarcoma. Exploiting matched RNA-seq for $N = 735$ of the PCAWG samples, we found that *BRCA1* and *BRCA2* have significantly lower expression in those samples with deletions at *BRCA1* and *BRCA2*, respectively, in comparison with those samples without a deletion ($\log_2$ fold change of no BRCA1 deletion vs. a BRCA1 deletion = 0.94; $P = 5.9 \times 10^{-12}$ and $\log_2$ fold change of no BRCA2 deletion vs. a BRCA2 deletion = 0.6; $P = 2.9 \times 10^{-6}$) after adjusting for primary site. Furthermore, *BRCA1* expression was consistently compromised within all cancer types with sufficient samples to test individually (**Fig. 6B** and **C**).

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

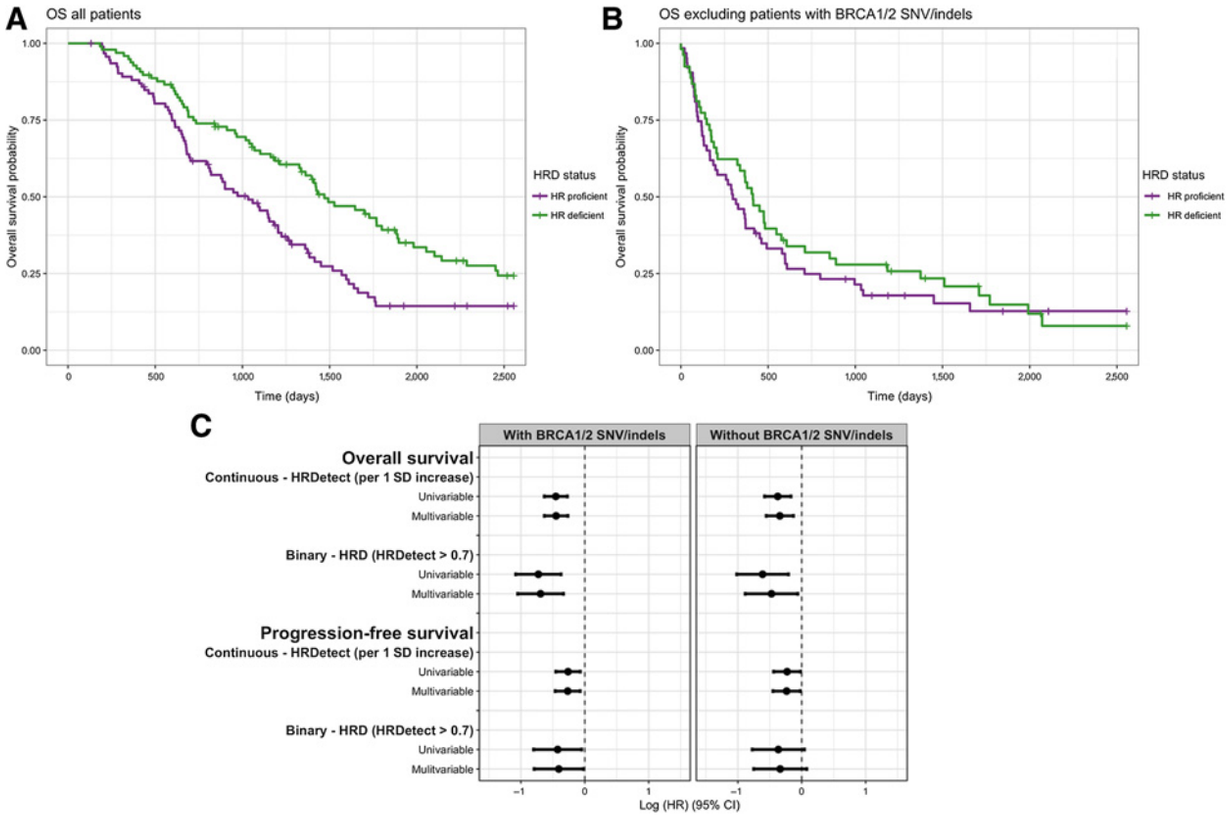

Figure 5.

Predicted HRD is associated with patient survival in the absence of short variants at *BRCA1/2*. **A,** The effect of HRD on overall survival time after diagnosis (in days) in HGSOC [*N* (events) = 190 (144)]. **B,** The effect of HRD on overall survival time after diagnosis (in days) in patients with HGSOC without *BRCA1/2* GSV/SSV [*N* (events) = 145 (113)]. **C,** Forest plot showing the effects of HRD on overall and PFS, unadjusted and adjusted for age and stage at diagnosis in a multivariable model. Estimates are also shown with tumors with SNVs/indels at *BRCA1/2* excluded. Kaplan–Meier plots compare survival times between HR-deficient and HR-proficient patients as defined by HRDetect score above and below 0.7. Hazard ratio estimates (on a log scale) are taken from Cox proportional hazards models and correspond to a 1 SD increase in HRDetect score, stratified by cohort and adjusted for age and stage at diagnosis.

Discussion

DNA repair deficiency in general and HRD in particular represent cancer cell vulnerabilities that have recently been exploited to exceptional patient benefit (14, 15, 17, 19–24, 46). However, the understanding of the genomic mechanisms that give rise to HRD is incomplete and the identification of patients whose tumors are HR deficient remains inaccurate.

In the largest collection of HGSOC WGS data examined to date, with matched expression data for most of the 205 tumors included, we have revealed new insights into the genesis of HRD in HGSOC based upon genome-wide mutational spectra. We show that structural variation at *BRCA1/2* in HGSOC is frequent and dominated by multi-megabase deletions, which are significantly enriched at *BRCA1* and *BRCA2* relative to the rest of the HGSOC genome. Examining transcriptomic data for the same samples, we have shown that large deletions overlapping *BRCA1/2* are associated with lower *BRCA1/2* expression, suggesting a direct impact of these deletions on gene function in many cases. Large deletions spanning *BRCA1/2* contribute to HRD independently of short variants, and samples with compound deletions affecting both *BRCA1* and *BRCA2* generate the highest risk of HRD. The frequent inactivation of *BRCA1/2* by large deletions in HGSOC is novel to our knowledge, and the original analysis of the AOCS cohort reported only one *BRCA1/2* large deletion (>1 Mb; ref. 5; Supplementary Table S4.2). The original TCGA cohort analysis did report frequent losses of the chr13q and chr17q chromosome arms (including the *BRCA1/2* loci) based upon SNP microarray data (ref. 6; Supplementary Table S5.1), but such data are known to generate high levels of false positives and negatives (47, 48) and these losses were not postulated to affect *BRCA1/2* function. Thus, previous assessments of SVs impacting the *BRCA1/2* loci have been characterized by underreporting, likely to be a result of the use of less sensitive algorithms tuned to detect smaller focal deletions (5), as well as copy-number alteration estimates derived from SNP microarray data (5, 6) and exome-restricted sequencing data (6, 49). Other types of structural variations are less frequent, but still evident, such as large inversions at *BRCA1* and duplications at *BRCA2*. The impact of these categories of mutation on the function of the gene is less well studied, but our data suggest that when *BRCA1* inversions in particular are considered together with the other mutational events in the tumor, their presence may aid prediction of HRD.

Our data also suggest a significant frequency of *BRCA1* and *BRCA2* loss by structural variation in a series of other cancers, including both cancers known to suffer *BRCA1* and *BRCA2* loss due to SNVs or indels, such as breast, pancreatic, and prostate cancer, as well as a number of

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

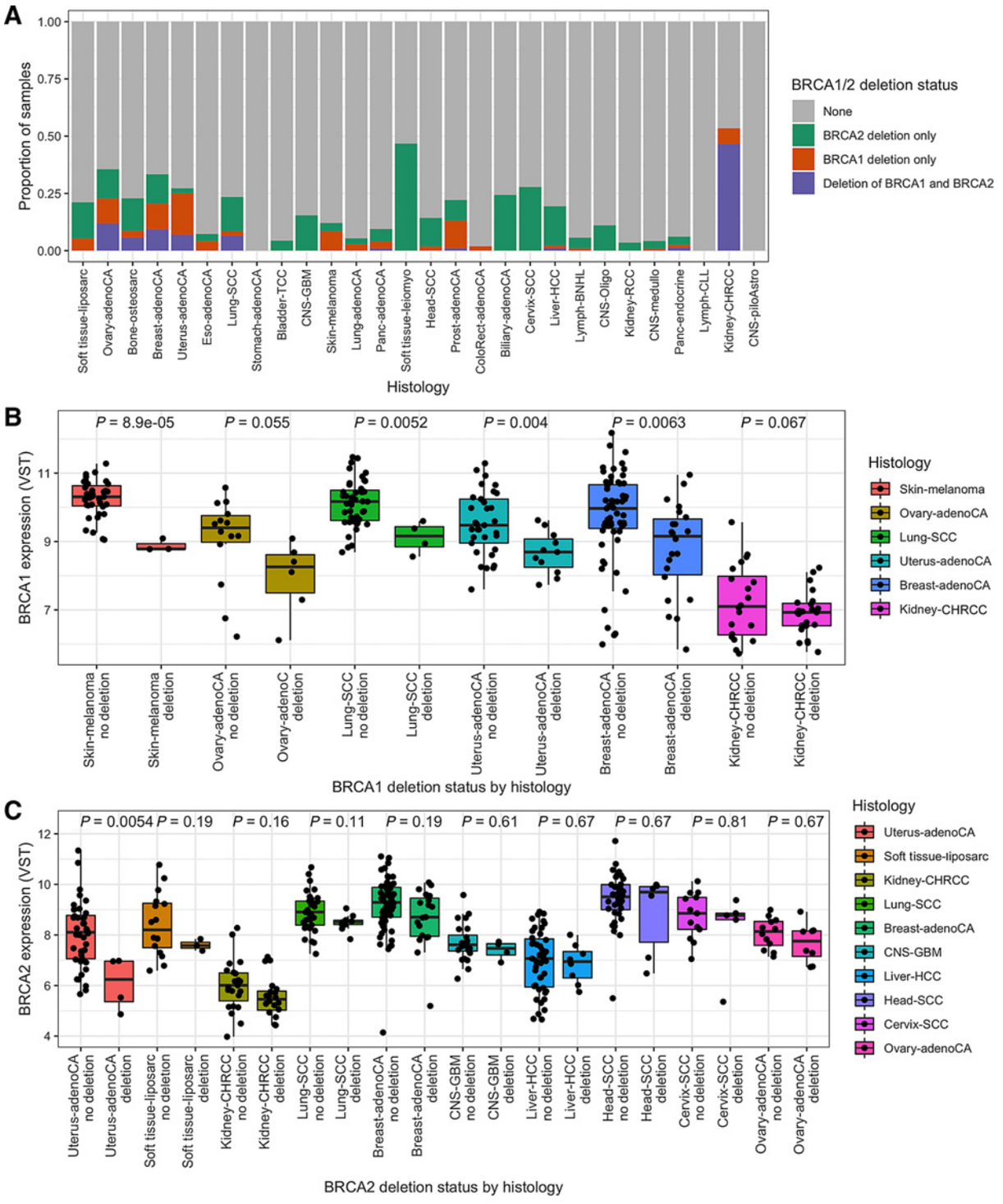

Figure 6.

Deletions at *BRCA1/2* in other cancer types and their impact on expression. **A,** The proportion of samples with deletions at *BRCA1, BRCA2*, or both, by primary site in PCAWG. **B,** *BRCA1* expression in tumors with and without deletions at *BRCA1* colored by primary site. **C,** *BRCA2* expression in tumors with and without deletions at *BRCA2* colored by primary site. Gene expression was visualized in variance stabilizing transformed (VST) counts. *P* values are from differential expression analyses conducted using DESeq2 for each primary site separately and are adjusted for multiple testing using the Benjamini–Hochberg approach. Only sites with at least 15 samples in total and at least three samples with a deletion at the gene in question were included.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

cancers outside of this group, including squamous cell lung cancer and cervical cancer. Perhaps most notably, soft-tissue leiomyosarcoma had a high incidence of large deletions at *BRCA2* (47%) and chromophobe renal cancers had a high incidence of large deletions of both *BRCA1* (53%) and *BRCA2* (47%). CHRCC often undergoes concurrent deletion of both *BRCA1* and *BRCA2* as a result of whole chromosome arm losses as has been reported previously (51), and trials of PARP inhibition as a therapy for renal cell carcinoma are ongoing (52). The fact that across these cancer types, large *BRCA1* and *BRCA2* deletions were associated with loss of gene expression suggests functionality is lost and that strategies to detect these genomic events and trials of PARP inhibition should also be considered in these patients.

Finally, we have constructed an integrated model of HRD in HGSOC, including a large variety of mutation- and expression-based variables across the combined cohort. This model supports an independent role for structural variation at *BRCA1/2* in HRD and highlights the diversity of routes that tumors may follow to reach HRD. Given this diversity, and the substantial fraction of samples where HRD is detected in the absence of any detectable *BRCA1/2* mutations, we conclude that the direct detection of HRD in HGSOC using genome-wide sequencing data is a valuable addition to the search for inactivating mutations in HR pathway genes. This is likely to be the case for other cancers showing evidence for HRD, such as uterine, lung squamous, esophageal, sarcoma, bladder, lung adenocarcinoma, head and neck, and gastric carcinomas (1). The variety of events sufficient for a tumor to develop HRD is not well understood, but recent studies suggest that there is selective pressure for biallelic inactivation leading to HRD in cancer types with predisposing germline variants in the HR pathway, such as breast, ovarian, pancreatic, and prostate cancers (53).

One of the key challenges in studies of this type is deciding upon a "gold standard" test of HRD. Current functional, clinical, and molecular tests all have advantages and disadvantages. The limitations of HRDetect that we used here include that it was developed and trained using breast tumor data and is predicated on BRCA1/2 deficiency arising from short variant disruption and promoter methylation, rather than any form of disruption to any HR gene. Although in the context of this study, of BRCA1/2 disruption by SV, the latter is of less importance. All current genomic HRD tests are further limited to demonstrating that HRD once existed in the evolution of a tumor, and are blind to the restoration of HR by events such as secondary mutations, hypomorphic HRD variants, and epigenomic changes.

There is an urgent clinical need to better understand the processes that give rise to both *BRCA1/2* loss and more broadly contribute to HRD. Our study demonstrates that *BRCA1/2* loss by structural variation may have a comparable impact on HRD and patient survival with short variants at *BRCA1/2*. Furthermore, these events are not specific to HGSOC, they are abundant, they compromise gene expression, and are likely to be functionally important in other cancer types, including in those types where HRD has been identified previously. However, these variants are unlikely to be detected by sequencing methods currently employed in the clinic. A change in sequencing approach to identify all *BRCA1/2* loss may be required to maximize the potential of PARP inhibition.

Authors' Disclosures

A. Ewing reports grants from UKRI during the conduct of the study. R.L. Hollis reports grants from AstraZeneca during the conduct of the study and personal fees from GlaxoSmithKline outside the submitted work. C.S. Herrington reports grants from AstraZeneca during the conduct of the study. N. McPhail reports personal fees from Eisai, Ipsen, and Bayer outside the submitted work. S. Dowson reports grants from Beatson Cancer Charity during the conduct of the study. R. Glasspool reports grants from AstraZeneca during the conduct of the study, as well as grants from Clovis Oncology and personal fees from AstraZeneca, Clovis Oncology, and GlaxoSmithKline outside the submitted work. B. McDade reports grants from Scottish Genomes Partnership during the conduct of the study. A. Matakidou reports other from AstraZeneca during the conduct of the study and GlaxoSmithKline outside the submitted work, and has AstraZeneca shares. B. Dougherty reports other from AstraZeneca during the conduct of the study and outside the submitted work. R. March reports employment with and is a shareholder of AstraZeneca; AstraZeneca is a pharmaceutical company, with commercial products in PARP inhibition. J.C. Barrett reports other from AstraZeneca during the conduct of the study and outside the submitted work. I.A. McNeish reports personal fees from Clovis Oncology, GlaxoSmithKline, Roche, Alkerrnes, and ScanCell and grants and personal fees from AstraZeneca outside the submitted work. A.V. Biankin reports other from Bristol Myers Squibb, AstraZeneca, MyTomorrows, Elstar Therapeutics, Cumulus Oncology, Modulus Oncology, Wollemia Oncology, ConcR, Cambridge Cancer Genomics, Agilent Technologies, Gabriel Precision Oncology, and Human.ai outside the submitted work. P. Roxburgh reports grants from AstraZeneca during the conduct of the study, as well as grants from AstraZeneca and personal fees from AstraZeneca and GlaxoSmithKline/Tesaro outside the submitted work. C. Gourley reports grants from AstraZeneca and Scottish Chief Scientist's Office during the conduct of the study. C. Gourley also reports grants from Novartis, Aprea, BerGenBio, and Medannexin; grants and personal fees from AstraZeneca, GlaxoSmithKline, Tesaro, MSD, Roche, and NuCana; and personal fees from Clovis, Foundation One, Chugai, Takeda, Sierra Oncology, and Cor2Ed outside the submitted work, as well as has a patent for PCT/US2012/040805 issued, PCT/GB2013/053202 pending, 1409479.1 pending, 1409476.7 pending, and 1409478.3 pending. No disclosures were reported by the other authors.

Authors' Contributions

A. Ewing: Conceptualization, data curation, formal analysis, investigation, visualization, methodology, writing–original draft, writing–review and editing. **A. Meynert:** Conceptualization, data curation, formal analysis, investigation, methodology, writing–review and editing. **M. Churchman:** Resources, data curation, project administration, writing–review and editing. **G.R. Grimes:** Data curation, formal analysis, investigation, methodology, writing–review and editing. **R.L. Hollis:** Resources, data curation, writing–review and editing. **C.S. Herrington:** Conceptualization, resources, data curation, writing–review and editing. **T. Rye:** Data curation, writing–review and editing. **C. Bartos:** Data curation, writing–review and editing. **I. Croy:** Resources, writing–review and editing. **M. Ferguson:** Conceptualization, resources, data curation, writing–review and editing. **M. Lennie:** Resources, data curation, writing–review and editing. **T. McGoldrick:** Conceptualization, resources, data curation, writing–review and editing. **N. McPhail:** Conceptualization, resources, data curation, writing–review and editing. **N. Siddiqui:** Resources, writing–review and editing. **S. Dowson:** Resources, data curation, writing–review and editing. **R. Glasspool:** Resources, data curation, writing–review and editing. **M. Mackean:** Data curation, writing–review and editing. **F. Nussey:** Resources, data curation, writing–review and editing. **B. McDade:** Resources, writing–review and editing, performed whole genome sequencing. **D. Ennis:** Resources, writing–review and editing. **L. McMahon:** Supervision, project administration, writing–review and editing. **A. Matakidou:** Conceptualization, supervision, writing–review and editing. **B. Dougherty:** Conceptualization, supervision, writing–review and editing. **R. March:** Conceptualization, supervision, writing–review and editing. **J.C. Barrett:** Conceptualization, supervision, writing–review and editing. **I.A. McNeish:** Conceptualization, supervision, writing–review and editing. **The Scottish Genomes Partnership:** Conceptualization, funding acquisition, writing–review and editing. **A.V. Biankin:** Conceptualization, supervision, funding acquisition, writing–review and editing. **P. Roxburgh:** Conceptualization, resources, data curation, supervision, investigation, writing–review and editing. **C. Gourley:** Conceptualization, resources, data curation, supervision, investigation, visualization, writing–original draft, writing–review and editing. **C.A. Semple:** Conceptualization, data curation, supervision, investigation, visualization, writing–original draft, writing–review and editing.

Acknowledgments

A. Ewing was supported by a UKRI Innovation Fellowship (MR/RO26017/1). C.A. Semple, A. Meynert, and G.R. Grimes were supported by MRC core funding to the MRC Human Genetics Unit (MRC grant MC_UU_00007/16). R.L. Hollis was supported by an MRC-funded Research Fellowship. S. Dowson received funding from AstraZeneca and the Beatson Cancer Charity. I.A. McNeish acknowledges funding from Ovarian Cancer Action and the NIHR Imperial Biomedical Research Centre. A.V. Biankin acknowledges funding from Cancer Research UK (C29717/A17263,

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

C29717/A18484, C596/A18076, C596/A20921, and A23526), Wellcome Trust Senior Investigator Award (103721/Z/14/Z), Pancreatic Cancer UK Future Research Leaders fund (FLF2015_04_Glasgow), MRC/EPSRC Glasgow Molecular Pathology Node, and The Howat Foundation. Sequencing of the SHGSOC cohort was supported by AstraZeneca, the Medical Research Council, and the Scottish Chief Scientist through a Precision Medicine Scotland Innovation Centre/Scottish Genome Partnership (SEHHD-CSO 1175759/2158447) collaboration. This Scottish Genomes Partnership was funded by the Chief Scientist Office of the Scottish Government Health Directorates (SGP/1) and The Medical Research Council Whole Genome Sequencing for Health and Wealth Initiative (MC/PC/15080). This study would not be possible without the families, patients, clinicians, nurses, research scientists, laboratory staff, informaticians, and the wider Scottish Genomes Partnership team to whom we give grateful thanks. Members of the Scottish Genome Partnership (SGP) include Timothy J. Aitman, Andrew V. Biankin, Susanna L. Cooke, Wendy Inglis Humphrey, Sancha Martin, Lynne Mennie, Alison Meynert, Zosia Miedzybrodzka, Fiona Murphy, Craig Nourse, Javier Santoyo-Lopez, Colin A. Semple, and Nicola Williams. More information about SGP can be found at www.scottishgenomespartnership.org. The authors would also like to acknowledge the Edinburgh Clinical Research Facility for the sequencing of RNA samples from the SHGSOC cohort. The authors would also like to extend our thanks to the Nicola Murray Foundation and the Edinburgh Ovarian Cancer Database, from which the clinical data for much of the Scottish cohort were retrieved. We also thank the NRS Lothian Human Annotated Bioresource, NHS Lothian Department of Pathology, the Edinburgh Experimental Cancer Medicine Centre, the Biorepository at the Glasgow Queen Elizabeth University Hospital, and the Tayside Biorepository for their support.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 20, 2020; revised January 29, 2021; accepted March 16, 2021; published first March 19, 2021.

References

1. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas. Cell Rep 2018;23:239–54.
2. Bowtell DD, Böhm S, Ahmed AA, Aspuria P-J, Bast RC Jr, Beral V, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nat Rev Cancer 2015;15:668–79.
3. 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.
4. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C, et al. Emerging landscape of oncogenic signatures across human cancers. Nat Genet 2013;45:1127–33.
5. 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.
6. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature 2011;474:609–15.
7. Hollis RL, Gourley C. Genetic and molecular changes in ovarian cancer. Cancer Biol Med 2016;13:236–47.
8. Chiang JW, Karlan BY, Cass L, Baldwin RL. BRCA1 promoter methylation predicts adverse ovarian cancer prognosis. Gynecol Oncol 2006; 101:403–10.
9. Hughes-Davies L, Huntsman D, Ruas M, Fuks F, Bye J, Chin S-F, et al. EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell 2003;115: 523–35.
10. Brown LA, Irving J, Parker R, Kim H, Press JZ, Longacre TA, et al. Amplification of EMSY, a novel oncogene on 11q13, in high grade ovarian surface epithelial carcinomas. Gynecol Oncol 2006;100:264–70.
11. Hollis RL, Churchman M, Michie CO, Rye T, Knight L, McCavigan A, et al. High EMSY expression defines a BRCA-like subgroup of high-grade serous ovarian carcinoma with prolonged survival and hypersensitivity to platinum. Cancer 2019;125:2772–81.
12. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. Cancer Discov 2015;5:1137–54.
13. Nguyen LWM, Martens J, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. Nat Commun 2020;11:1–12.
14. Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: A preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. Lancet Oncol 2014;15:852–61.
15. Mirza MR, Monk BJ, Herrstedt J, Oza AM, Mahner S, Redondo A, et al. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. N Engl J Med 2016;375:2154–64.
16. Coleman RL, Oza AM, Lorusso D, Aghajanian C, Oaknin A, Dean A, et al. Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;390:1949–61.
17. Moore K, Colombo N, Scambia G, Kim B-G, Oaknin A, Friedlander M, et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. N Engl J Med 2018;379:2495–505.
18. Ray-Coquard I, Pautier P, Pignata S, Pérol D, González-Martín A, Berger R, et al. Olaparib plus bevacizumab as first-line maintenance in ovarian cancer. N Engl J Med 2019;381:2416–28.
19. González-Martín A, Pothuri B, Vergote I, Christensen RD, Graybill W, Mirza MR, et al. Niraparib in patients with newly diagnosed advanced ovarian cancer. N Engl J Med 2019;381:2391–402.
20. Robson M, Im S, Senkus E, Xu B, Domchek S, Masuda N, et al. Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. N Engl J Med 2017;377:523–33.
21. Litton JK, Rugo HS, Ettl J, Hurvitz SA, Gonçalves A, Lee K-H, et al. Talazoparib in patients with advanced breast cancer and a germline BRCA mutation. N Engl J Med 2018;379:753–63.
22. Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ, et al. Maintenance olaparib for germline BRCA-mutated metastatic pancreatic cancer. N Engl J Med 2019;381:317–27.
23. de Bono J, Mateo J, Fizazi K, Saad F, Shore N, Sandhu S, et al. Olaparib for metastatic castration-resistant prostate cancer. N Engl J Med 2020;382: 2091–102.
24. Gourley C, Balmaña J, Ledermann JA, Serra V, Dent R, Loibl S, et al. Moving from poly (ADP-Ribose) polymerase inhibition to targeting DNA repair and DNA damage response in cancer therapy. J Clin Oncol 2019;37:2257–69.
25. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature 1995:378;789–92.
26. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994;266:66–71.
27. Paluch-Shimon S, Cardoso F, Sessa C, Balmana J, Cardoso MJ, Gilbert F, et al. Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO Clinical Practice Guidelines for cancer prevention and screening. Ann Oncol 2016;27:v103–v110.
28. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer 2016;16:110–20.
29. Guan P. Structural variation detection using next-generation sequencing data: a comparative technical review. Methods 2016;102:36–49.
30. Ewing A, Semple C. Breaking point: the genesis and impact of structural variation in tumours. F1000Research 2018;7:1814.
31. GitHub - bcbio/bcbio-nextgen: validated, scalable, community developed variant calling, RNA-seq and small RNA analysis. Available from: https://github.com/bcbio/bcbio-nextgen.
32. Landrum MJ, Kattman BL. ClinVar at five years: delivering on the promise. Hum Mutat 2018;39:1623–30.
33. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019;47:D886–94.
34. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32:1220–2.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022

35. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-Cancer Analysis of Whole Genomes. Nature 2020;578:82–93.
36. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R, et al. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 2016;32: 289–91.
37. Zhao EY, Shen Y, Pleasance E, Kasaian K, Leelakumari S, Jones M, et al. Homologous recombination deficiency and platinum-based therapy outcomes in advanced breast cancer. Clin Cancer Res 2017;23: 7521–30.
38. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14:417–9.
39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
40. Irodi A, Rye T, Herbert K, Churchman M, Bartos C, Mackean M, et al. Patterns of clinicopathological features and outcome in epithelial ovarian cancer patients: 35 years of prospectively collected data. BJOG 2020;127: 1409–20.
41. Dougherty BA, Lai Z, Hodgson DR, Orr MCM, Hawryluk M, Sun J, et al. Biological and clinical evidence for somatic mutations in BRCA1 and BRCA2 as predictive markers for olaparib response in high-grade serous ovarian cancers in the maintenance setting. Oncotarget 2017;8:43653–61.
42. 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.
43. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68:820–3.
44. Knudson AG. Two genetic hits (more or less) to cancer. Nat Rev Cancer 2001;1: 157–62.
45. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole genome analyses of metastatic solid tumors. Nature 2019; 575:210–6.
46. Swisher EM, Lin KK, Oza AM, Scott CL, Giordano H, Sun J, et al. Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. Lancet Oncol 2017;18:75–87.
47. Haraksingh RR, Abyzov A, Urban AE. Comprehensive performance comparison of high-resolution array platforms for genome-wide copy number variation (CNV) analysis in humans. BMC Genomics 2017;18:321.
48. Zhou B, Ho SS, Zhang X, Pattni R, Haraksingh RR, Urban AE, et al. Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. J Med Genet 2018; 55:735–43.
49. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A comprehensive pan-Cancer molecular study of gynecologic and breast cancers. Cancer Cell 2018;33:690.
50. Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of somatic structural variation in human cancer genomes. Nature 2020;578: 112–21.
51. Rathmell KW, Chen F, Creighton CJ. Genomics of chromophobe renal cell carcinoma: implications from a rare tumor for pan-cancer studies. Oncoscience 2015;2:81–90.
52. Pilié PG. Genomic instability in kidney cancer: etiologies and treatment opportunities. Kidney Cancer 2019;3:143–50.
53. Jonsson P, Bandlamudi C, Cheng ML, Srinivasan P, Chavan SS, Friedman ND, et al. Tumour lineage shapes BRCA-mediated phenotypes. Nature 2019;571:576–9.

Downloaded from http://aacrjournals.org/clincancerres/article-pdf/27/11/3201/3088524/3201.pdf by guest on 27 August 2022