

# Aberrant Transcript Usage Is Associated with Homologous Recombination Deficiency and Predicts Therapeutic Response



Hyeon Gu Kang<sup>1</sup>, Haeun Hwangbo<sup>1</sup>, Myung Ji Kim<sup>2</sup>, Sinae Kim<sup>2</sup>, Eun Ji Lee<sup>2</sup>, Min Ji Park<sup>2</sup>, Jae-Weon Kim<sup>3</sup>, Byoung-Gie Kim<sup>4</sup>, Eun-Hae Cho<sup>5</sup>, Suhwan Chang<sup>2</sup>, Jung-Yun Lee<sup>6</sup>, and Jung Kyoon Choi<sup>1,7</sup>

## ABSTRACT

BRCA1/2 mutations account for only a small fraction of homologous recombination (HR) deficiency (HRD) cases. Recently developed genomic HRD (gHRD) tests suffer confounding factors that cause low precision in predicting samples that will respond to PARP inhibitors and DNA damaging agents. Here we present molecular and clinical evidence of transcriptional HRD (tHRD) that is based on aberrant transcript usage (aTU) of minor isoforms. Specifically, increased TU of nonfunctional isoforms of DNA repair genes was prevalent in breast and ovarian cancer with gHRD. Functional assays validated the association of aTU with impaired HR activity. Machine learning-based tHRD detection by the transcript usage (TU) pattern of key genes was superior to directly screening for

gHRD or BRCA1/2 mutations in accurately predicting responses of cell lines and patients with cancer to PARP inhibitors and genotoxic drugs. This approach demonstrated the capability of tHRD status to reflect functional HR status, including in a cohort of olaparib-treated ovarian cancer with acquired platinum resistance. Diagnostic tests based on tHRD are expected to broaden the clinical utility of PARP inhibitors.

**Significance:** A novel but widespread transcriptional mechanism by which homologous recombination deficiency arises independently of BRCA1/2 mutations can be utilized as a companion diagnostic for PARP inhibitors.

## Introduction

The two key genes of the homologous recombination (HR) pathway for double-strand break (DSB) repair, BRCA1 and BRCA2, are intimately associated with the heritability of breast and ovarian cancer. Recent studies have shown that DNA damage response (DDR) pathways are frequently mutated in these and other cancers (1–3). Therefore, targeting vulnerabilities in DNA repair activity, in particular HR

deficiency (HRD), has been a primary treatment option. In the representative example of synthetic lethality, inhibiting PARP enzymes can selectively suppress cells that are defective in DDRs, for example, due to BRCA1/2 mutations (4–9).

However, BRCA1/2 mutations account for a small fraction of HRD cases, thereby limiting the development of more comprehensive HRD diagnostic markers. A promising approach has been to detect the genomic consequences, instead of the causal factors, of HRD. Defective HR machinery leaves genomic footprints and lesions as a consequence of failure in damage maintenance. Specifically, mutational signature 3 and genomic scar are well known indicators of HRD (10–15). Signature 3 has been linked to HRD by associating the genome-wide patterns of single-nucleotide variants with specific background factors (16). Genomic scar is determined by a combination of three chromosomal aberrant events, namely, telomeric allelic imbalances, large-scale state transition, and LOH (17–19).

These genomic assays, however, cannot reflect the functional restoration of HR that gives rise to therapeutic resistance (20). BRCA1/2 reversion and various other mechanisms can account for acquired resistance to platinum regimens and PARP inhibitors in ovarian cancer (21–27). In these cases, initially imprinted genomic scars remain detectable in tumors with recovered HR activity. Furthermore, several technical artifacts can arise when calculating mutational signatures (28). For example, the bleeding of signatures refers to the phenomenon in which signatures present in only some samples of a cohort affect the signature assignment of the entire cohort. Directly related to detection of HRD is the tendency of the fitting processes to force the assignment of flat signature 3 to irrelevant samples (28).

These biological or technical confounding factors together make detection of genomic HRD (gHRD) prone to false positives. This results in low levels of precision because a considerable number of tumors predicted to respond to DNA-damaging agents or PARP inhibitors would not actually respond (29). Because of the coupling

<sup>1</sup>Department of Bio and Brain Engineering, KAIST, Daejeon, Republic of Korea. <sup>2</sup>Department of Biomedical Engineering, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. <sup>3</sup>Department of Obstetrics and Gynecology, Seoul National University Hospital, Seoul, Republic of Korea. <sup>4</sup>Department of Obstetrics and Gynecology, Gynecologic Cancer Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. <sup>5</sup>Genome Research Center, GC Genome, GC Labs, Yongin-si, Gyeonggi-do, Republic of Korea. <sup>6</sup>Department of Obstetrics and Gynecology, Yonsei University College of Medicine, Seoul, Republic of Korea. <sup>7</sup>PentaMedix Co., Ltd., Seongnam-si, Gyeonggi-do, Republic of Korea.

**Corresponding Authors:** Eun-Hae Cho, GC Genome, GC Labs, 107 Ihyeonro, 30beon-gil, Giheung-gu, Yongin-si, Gyeonggi-do 16924, Republic of Korea. Phone: 82-31-260-9216; E-mail: ehcho@gccorp.com; Suhwan Chang, Department of Biomedical Engineering, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Pungnap 2(i)-dong, Songpa-gu, Seoul, Republic of Korea. Phone: 82-2-3010-2095; E-mail: suhwan.chang@amc.seoul.kr; Jung-Yun Lee, Department of Obstetrics and Gynecology, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea. Phone: 82-2-2228-2237; E-mail: jungyunlee@yuhs.ac; and Jung Kyoon Choi, YBS Building (E16-1), KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. Phone: 82-42-350-4327; E-mail: jungkyoon@kaist.ac.kr

Cancer Res 2022;82:142–54

doi: 10.1158/0008-5472.CAN-21-2023

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2021 The Authors; Published by the American Association for Cancer Research

of precision and recall, attempts to improve precision by applying more stringent cut-off points will undermine sensitivity. This calls for a different biomarker that can complement this limitation and improve precision. For example, biomarkers that can detect functional HRD status may be used to refine gHRD predictions by filtering out tumors that regained HR functionality (29).

In this work, we pay attention to transcriptional processes, which change dynamically in response to functional fluctuation during cancer evolution but are not affected significantly by the genomic consequences of HRD. We employ machine learning to capture patterns of transcriptional aberration that may be responsible for true gHRD readouts by contrasting with the profiles of negative gHRD samples. Ultimately, by leveraging data from patients with acquired resistance to initial treatments, we attempt to test whether our method can determine real-time, functional HR status and overcome the limitations imposed by the artefacts of the gHRD tests.

## Materials and Methods

### The Cancer Genome Atlas data collection and gHRD determination

Of 757 cases with breast adenocarcinoma from The Cancer Genome Atlas (TCGA; TCGA-BRCA), we identified 644 samples with the required molecular data available for our analysis, including bam files from RNA sequencing (RNA-seq). We analyzed 324 TCGA ovarian serous cystadenocarcinoma samples (TCGA-OV) whose chemotherapy response data were available. Processing of the data is elaborated in Supplemental Methods.

To define differences in expression level, splicing, and transcript usage (TU) between HRD<sup>+</sup> and HRD<sup>-</sup>, we used a stringent classification of gHRD based on using both genomic scar and signature 3. Genomic scar (HRD scores) and mutation signature 3 were obtained from TCGA Hub (<https://tcga.xenahubs.net>) and mSignatureDB (<http://tardis.cgu.edu.tw/msignaturedb>). Six hundred and forty-four TCGA-BRCA samples were marked as gHRD<sup>+</sup> when both genomic scar and signature 3 exceeded the median level ( $n = 210$ ). Samples were marked as gHRD<sup>-</sup> when both genomic scar and signature 3 were less than the median level ( $n = 225$ ). These 435 samples were retained for further analysis. We analyzed 324 TCGA-OV samples in the same manner. The samples were marked as gHRD<sup>+</sup> when both genomic scar and signature 3 exceeded the median level ( $n = 80$ ). Samples were marked as gHRD<sup>-</sup> when both genomic scar and signature 3 were less than the median level ( $n = 82$ ).

### Quantification of TU and identification of aberrant TU

Individual transcript levels were quantified by transcripts per million (TPM) through the two-pass method of StringTie (v.1.3.5; ref. 30) using recommended parameters. For quantification of all transcripts, including novel transcripts, we first performed the assembly step using GENCODE v29 annotation for the BRCA1/2-active TCGA-BRCA samples ( $n = 388$ ) and TCGA-OV samples ( $n = 324$ ), and GENCODE v19 annotation for the Cancer Cell Line Encyclopedia (CCLE) samples ( $n = 56$ ) as reference data, to extract annotated and novel transcripts as individual gene transfer format (GTF) files for each sample. We then created a newly assembled reference by merging these GTF files of our different samples, which were used for transcript quantification. TU was calculated by first matching the gene origin of all transcripts using GffCompare (v.0.11.2; 31) and then calculating the ratio of the respective TPM transcript to the sum of all TPM transcripts belonging to the respective matched gene.

Additional filtering was performed to identify minor transcripts with functional loss from the computed TU matrix. Major transcripts encoding proteins that are well preserved functionally and evolutionarily (defined as principal isoforms; Principals 1–5) were filtered using GENCODE v29 annotation from the Annotating Principal Splice Isoforms (APRIS) database (32). The remaining transcripts that passed through this initial filter belonged to the Alternative 1, Alternative 2, or Not-Reported category, and were subsequently classified as minor transcripts for our use in further analysis. Next, the mean counts per million (CPM) of each gene was calculated to reduce noise in the minor TU matrix; the TU of genes with a mean CPM less than 1 was filtered out. Additionally, standard deviation-based filtering was performed to remove TU with low intersample differences and to minimize noise in the minor TU matrix.

We used two reciprocal approaches to identifying aberrant TU (aTU) events. First, we identified minor transcripts whose TU was significantly different between gHRD<sup>+</sup> and gHRD<sup>-</sup> samples by using the Mann–Whitney  $U$  test. The  $P$  values were adjusted based on Benjamini–Hochberg correction for TCGA-BRCA but not TCGA-OV because of small sample size. We selected the minor transcript with the highest statistical significance among multiple minor transcripts from the same gene showing significant intergroup differences (FDR < 1% for TCGA-BRCA and nominal  $P < 0.01$  for TCGA-OV). aTU in gHRD<sup>+</sup> was defined when the most significant minor transcript was overexpressed in gHRD<sup>+</sup>. On the other hand, aTU in gHRD<sup>-</sup> was defined when the most significant minor transcript was overexpressed in gHRD<sup>-</sup>. Second, we partitioned samples based on TU values instead of gHRD status. For each minor transcript, we selected samples whose TU was greater than the 75<sup>th</sup> percentile ('high minor TU' group) and those whose TU was lower than the 25<sup>th</sup> percentile ('low minor TU' group). We then compared the genomic scar scores between the two groups using the Mann–Whitney  $U$  test followed by Benjamini–Hochberg correction for TCGA-BRCA. This procedure was repeated for each minor transcript from respective genes to identify significant differences in genomic scar (FDR < 10% for TCGA-BRCA and nominal  $P < 0.01$  for TCGA-OV). The transcript with the most significant difference in genomic scar was chosen for each gene. aTU in gHRD<sup>+</sup> was defined when genomic scar was higher in the 'high minor TU' group for the most significant transcript. On the contrary, aTU in gHRD<sup>-</sup> was defined when genomic scar was higher in the 'low minor TU' group for the most significant transcript.

### Development of the tHRD prediction model

For the breast cancer model, we used all minor transcripts ( $n = 104$ ) of DNA repair genes with aTU in gHRD<sup>+</sup> ( $n = 35$ ) as input to the random forest (RF) classifier. TCGA-BRCA samples ( $n = 388$ ) were divided in a 7:3 ratio, representing the training set ( $n = 272$ ) and test set ( $n = 116$ ) respectively, and the TU of the 104 isoforms were used as features for prediction of gHRD status. Best parameter tuning of the RF classifier was performed using the RandomizedSearchCV module of sklearn.model. Each set of parameters went through 100 iterations, and a stratified three-fold cross-validation was performed. Mean validation accuracy was computed for each result parameter set. Parameter sets were sorted based on this mean value to identify the parameter set with the highest mean validation accuracy. The average validation accuracy for the classifier model developed with the best parameter set was 0.82. The accuracy of the model for the test set indicated 0.84. The model developed by this process was applied to the CCLE dataset to analyze the correlations of tHRD status with sensitivity to PARP inhibitors or DNA damaging drugs with the binary classification of tHRD<sup>+</sup> or tHRD<sup>-</sup> using a threshold score of 0.5. The tHRD probability itself was

used for its correlation with gene dependency (DepMap RNAi combined score).

The same procedures were repeated for the ovarian cancer model with all minor forms ( $n = 74$ ) of DNA repair genes with aTU in gHRD<sup>+</sup> genes ( $n = 22$ ). The 74 TU values were used as features to develop the RF classifier model. TCGA-OV samples were divided into a training set ( $n = 130$ ) and test set ( $n = 32$ ). Because of its smaller sample size compared with TCGA-BRCA set, we performed a stricter best parameter tuning of the ovarian classifier to prevent overfitting. This tuning was performed by the ten-fold cross-validation option of RandomizedSearchCV. The parameter set with the highest mean validation accuracy was chosen as the best one. The accuracy of the model for the test set indicated 0.72. The model developed by this process was applied to predict platinum chemotherapy sensitivity. We employed the reannotation of the therapeutic responses of TCGA-OV samples to platinum chemotherapy (33). Model performance was evaluated by the precision-recall curve in comparison to the gHRD metrics. The identical model was subsequently used to predict therapeutic responses to olaparib and platinum in our neoadjuvant chemotherapy (NAC), platinum resistance (PR), olaparib maintenance (OM), and olaparib salvage (OS) cohorts.

### Clinical data collection and analysis

The NAC, PR, OM, and OS cohorts consisted of patients with advanced-stage or recurrent ovarian cancer who were treated with platinum-based chemotherapy and/or olaparib from 2016 to 2021 at 3 hospitals (Severance Hospital, Samsung Medical Center, and Seoul National University Hospital). The NAC cohort ( $n = 27$ ) included patients who were treated with NAC followed by interval debulking surgery. Fresh-frozen tumor samples were obtained before platinum-based chemotherapy. The PR cohort ( $n = 36$ ) represented cases that showed recurrence after platinum-based chemotherapy. Fresh-frozen tumor samples were obtained after acquisition of PR (at progression). Platinum responses were regarded as sensitive with platinum-free interval (PFI)  $\geq 6$  months or resistant with PFI  $< 6$  months. In the NAC cohort, 14 patients were classified as platinum-sensitive and 6 patients were platinum-resistant. All patients from the PR cohort were considered as platinum-resistant. In the OM ( $n = 24$ ) and OS ( $n = 33$ ) cohorts, patients with advanced-stage ovarian cancer treated with olaparib were enrolled for this study. The OS cohort represented patients who received olaparib as salvage therapy with PR and at least three prior lines of treatment. Most of the patients were BRCA1/2 mutants. Tumor samples, obtained before olaparib treatment, were embedded in paraffin after formalin fixation or kept fresh. The OM cohort included patients treated with olaparib for maintenance following platinum-based chemotherapy. Most of the patients were BRCA1/2 mutants. Tumor samples, obtained before the last round of platinum therapy, were embedded in paraffin after formalin fixation of kept fresh. Different criteria were applied to the OM and OS cohort when classifying the olaparib users into responders and nonresponders. Complete response or partial response was considered as durable response for the OS cohort. For the OM cohort, we considered patients with progression-free survival (PFS)  $\geq 12$  months as responders, based on Study 19 (ClinicalTrials.gov identifier: NCT00753545) reporting the median PFS of 11.2 months for patients with OM with BRCA1/2 mutations (34).

Clinical response was evaluated by the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1. PFS was calculated from the start date of therapy to the date of progression or death. This study was approved by the institutional review board of the participating centers (Severance Hospital, Samsung Medical Center, and Seoul

National University Hospital) in accordance with the Declaration of Helsinki and the International Conference of Harmonisation Good Clinical Practice Guidelines (Severance Hospital, 4-2018-0342; Samsung Medical Center, 2018-10-009 and 2019-03-126; Seoul National University Hospital, 1810-035-977). Informed written consent was obtained from all patients enrolled in this study. We used a panel (CancerSCAN 3.1) to determine the presence or absence of pathogenic or likely pathogenic mutations in 15 HR repair genes (i.e., BRCA1, BRCA2, ATM, BRIP1, PALB2, RAD51C, BARD1, CDK12, CHEK1, CHEK2, FANCL, PPP2R2A, RAD51B, RAD51D, and RAD54L). Classification of variants was conducted in accordance with the principles published in the American College of Medical Genetics and Genomic standards and guidelines for interpretation of sequence variants (35).

### Sequencing and gHRD estimation of cohort samples

Whole-genome sequencing (WGS), whole-exome sequencing (WES), and RNA-seq data were obtained from fresh-frozen or formalin-fixed, paraffin-embedded (FFPE) samples. Detailed sequencing procedures are described in Supplemental Methods.

To estimate the signature 3 contribution of the PR, NAC, OS, and OM samples, single-nucleotide variant (SNV) calling was performed using Mutect2 (v.2.2) with recommended filtering steps and default parameters. For the PR and OS cohort, Mutect2 matched-normal mode with default parameters was used with WES bam files as input for SNV calling, whereas Mutect2 tumor-only mode with default parameters was used for SNV calling from WGS bam files as input for the NAC and OM cohort. For the CCLE samples, we obtained the calling data for somatic mutations from the Broad Institute CCLE portal (<https://portals.broadinstitute.org/ccle>). Next, we determined the mutation signature of each SNV using the R package deconstructSigs (v.1.8.0; ref. 36). Variant information in the vcf files was transformed into the recommended input file format by using the vcf.to.sigs input method provided by the software package. To assess signature contributions, fitting was performed for the previously reported set of mutational signatures for each cancer type (i.e., signatures 1, 3, and 5 in ovarian cancer and signatures 1, 2, 3, 5, 6, 8, 10, 13, 17, 18, 20, 26, and 30 in breast cancer) using the Catalogue of Somatic Mutations in Cancer (COSMIC) database ([https://cancer.sanger.ac.uk/cosmic/signatures\\_v2](https://cancer.sanger.ac.uk/cosmic/signatures_v2)).

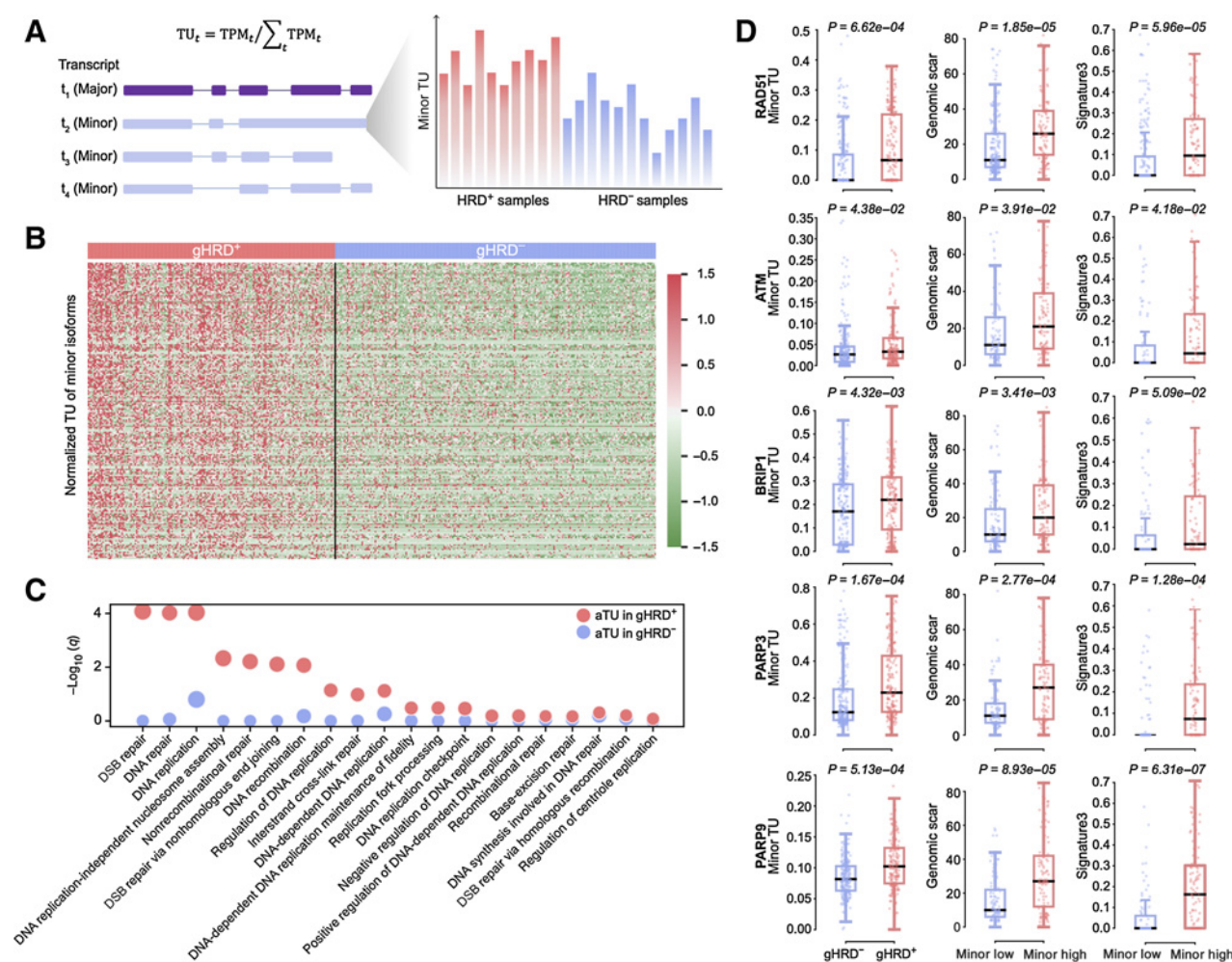
Genomic scar was estimated by determining copy number alterations in the WES or WGS data using sequenza-utils (v.3.0.0; ref. 37), based on which LOH, large scale transitions, and the number of telomeric allelic imbalances were estimated using the scarHRD R package (R package v.0.1.1; ref. 38). The sum of these values served as the genomic scar score (17–19). We calculated genome-wide guanine-cytosine (GC) content values by using the gc\_wiggle program with recommended parameters. We then processed the genomic bam files with the GC content values by using the bam2seqz program with recommended parameters to convert to a seqz file that contains information on copy number alterations. The process to determine copy-number alterations were identical to SNV calling with Mutect2. The matched-normal mode was used for the PR and OS WES bam files. The tumor-only mode was used for the NAC WGS bam files with the Pan-Cancer Analysis of Whole Genomes normal WGS bam file as nonmatching normal samples. Because the seqz file of each cohort obtained through sequenza-utils was split for each chromosome, a single seqz file was merged per sample. Genomic scar was calculated using the scarHRD R package (R package v0.1.1; ref. 38) and used as the input seqz file. gHRD status was determined by using the median of signature 3 and genomic scar scores in respective cohorts.

**Functional assays**

Primary breast cancer (PD:tHRD<sup>+</sup> and PD:tHRD<sup>-</sup>) cells were isolated from tumor tissues of a patient-derived xenograft (PDX) model. Fresh tumor tissues were minced into 1 to 2 mm pieces using sterile scissors, scalpel, and forceps. For tissue digestion, the tissue pieces were incubated in RPMI 1640 medium (Hyclone) containing 5% FBS (Hyclone), 1% penicillin/streptomycin (Hyclone), 20 µg/mL of collagenase Type III (Sigma Aldrich), and 840 ng/mL Amphotericin B (Gibco) for 6 hrs at 37°C. The digested tissue pieces were washed with media three times. The primary cells were cultured in RPMI 1640 medium containing 5% FBS, 1% penicillin/streptomycin, hEGF (20 ng/mL; Gibco), hydrocortisone (4 µg/mL; Sigma Aldrich), and transferrin (4 µg/mL; Sigma Aldrich). Depletion of mouse stromal cells was performed using Mouse Depletion Kit (Miltenybiotec) according to the manufacturer’s instructions.

For immunocytochemistry, the PD:tHRD<sup>+</sup> and PD:tHRD<sup>-</sup> cells were cultured in a slide-covered dish and treated with Olaparib (2 µmol/L; Sigma) for 48 hours. The cells were fixed with 4% formaldehyde and permeabilized with 0.1% Triton X-100. Sample slides were labeled with primary antibodies for Rad51 (1:500; Abcam) or γH2AX (1:1000; Abcam) and then stained with the corresponding Alexa Fluor 488-conjugated secondary antibody (Invitrogen) for 1 hour at room temperature. DAPI was used for nuclear staining. The cells were then mounted with Aqua-Poly-Mount mounting medium and imaged using an LSM710 confocal microscope (Carl Zeiss AG) under 400X magnification. Twenty cells from each sample were counted to calculate the nuclear Rad51 and γH2AX foci.

For HR activity assays, the PD:tHRD<sup>+</sup> and PD:tHRD<sup>-</sup> cells were pretreated with control media or with Olaparib (2 µmol/L). After 12 to



**Figure 1.** HRD-associated overexpression of minor transcripts of DNA repair genes. **A**, Concept of aTU as the relative overexpression of a minor isoform of a gene in HRD<sup>+</sup> tumors. TU is calculated as the TPM of an isoform divided by the sum of all TPMs. From each gene, the minor isoform with the most significant differential TU was selected and used to calculate the representative minor TU of the given gene. **B**, Normalized minor TU levels of 2,521 aTU genes in gHRD<sup>+</sup> versus gHRD<sup>-</sup> TCGA breast tumors. Significantly differential TU was identified at a FDR < 10%. **C**, Functional enrichment analysis of the 2,521 aTU genes (red dots) in comparison to a negative control set, for which aTU was defined as the relative overexpression of a minor isoform in gHRD<sup>-</sup> (blue dots). **D**, Representative aTU genes involved in DSB repair. Shown are the levels of the minor TU of the gene in gHRD<sup>+</sup> versus gHRD<sup>-</sup> samples (left), genomic scar in samples with high minor TU versus low minor TU (middle), and signature 3 in samples with high minor TU versus low minor TU (right). High and low minor TU are defined at the 75<sup>th</sup> and 25<sup>th</sup> percentile of all samples, respectively. Comparisons between two groups were performed using the Mann-Whitney *U* test, and the *P* values were corrected by the Benjamini-Hochberg procedure.

16 hrs, the cells were transfected with circular pDR-GFP (Addgene 26475) using Lipofectamine 2000 as recommended by the manufacturer (Invitrogen). After 24 hours, the cells were additionally transfected with the I-SceI-producing plasmid pCBASceI (Addgene 26477). Flow cytometry analysis was performed 24 hours later with AccuriFlow Cytometry (BD Biosciences). The GFP-positive cells were quantified by the FlowJo program, and the average HR frequency was measured from three independent samples.

We performed Western blotting as follows. After Olaparib treatment for 48 hours, protein samples were prepared and loaded at 10 to 50  $\mu$ g of total protein per lane. The blot was probed with anti-BCL2 (1:1000, Cell Signaling Technology), anti-BAX (1:1000, Cell Signaling Technology), anti-phospho-p53 (Ser-15 or -392; 1:1000; Cell Signaling Technology), and anti- $\beta$ -actin (1:1000; Santa Cruz Biotechnology) antibodies. The relative densities of bands were analyzed with the NIH ImageJ 1.47v software. For the alamarBlue cell viability assay, cells were cultured in a 96-well plate and treated with 2 mmol/L Olaparib. At 24, 48, and 72 hours

posttreatment, 1/10<sup>th</sup> volume of alamarBlue reagent (Invitrogen) was added directly to the culture media. The alamarBlue assay was performed according to the manufacturer's instructions. Viable cells were detected by fluorescence measurements using a microplate fluorescence spectrophotometer (GenTeks Biosciences, Inc.).

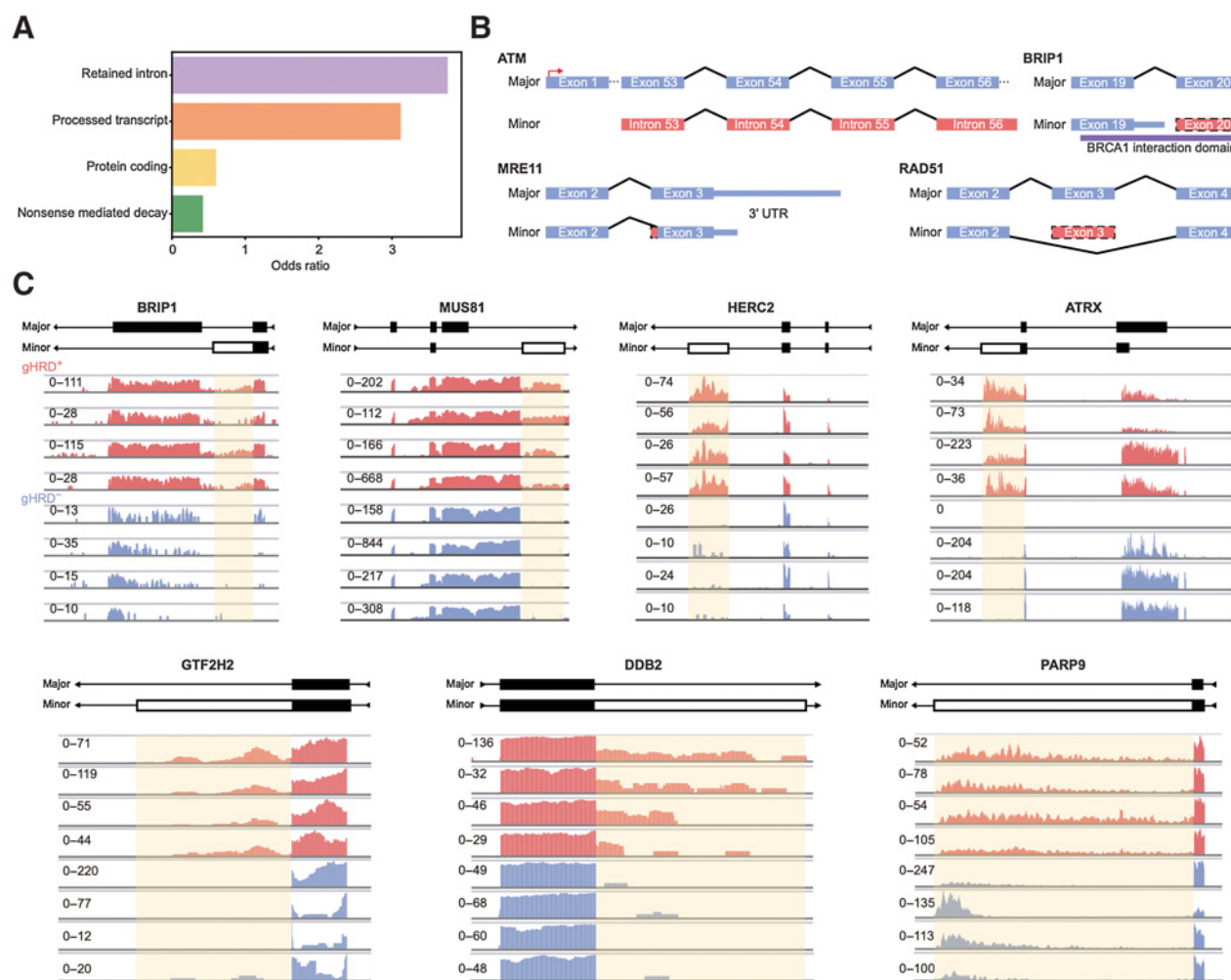
#### Data availability

Raw sequencing data for our cohort samples have been submitted to the BioProject database under BioProject ID PRJNA700673 (reviewer access: <http://www.ncbi.nlm.nih.gov/bioproject/700673>).

## Results

### Minor isoform overexpression is associated with gHRD

We collected breast cancer samples from TCGA, examined germline and somatic BRCA1/2 mutations, and defined HR status by using both genomic scar and signature 3 readouts (Supplementary Fig. S1A).



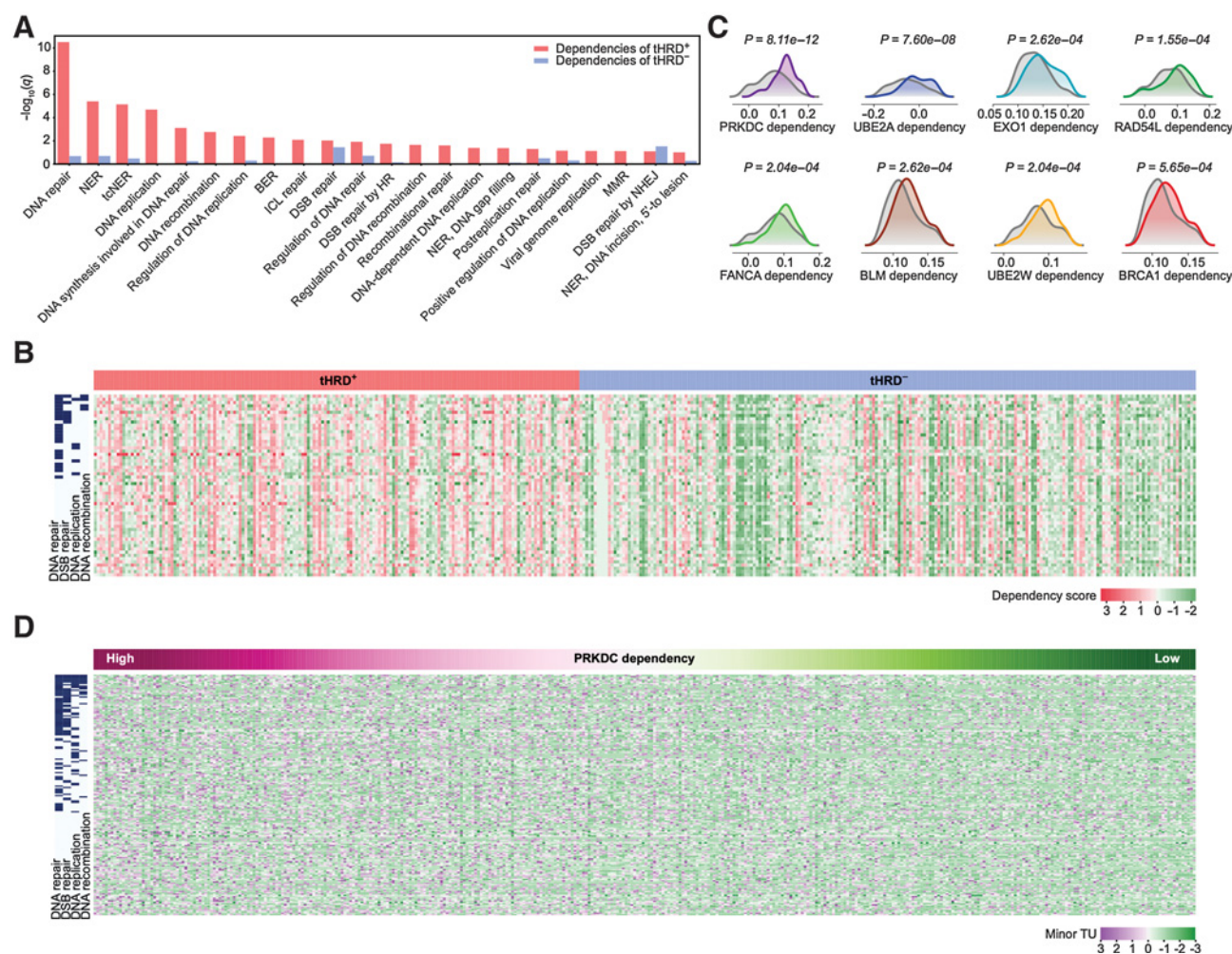
**Figure 2.**

Functional characterization of aTU events. **A**, Frequencies of annotation categories that represent the functional consequences or alternative events by minor isoform transcription. The odds of each class among the identified aTU genes was divided by the odds of that class among all genes. For example, a >3-fold overrepresentation of intron retention and noncoding transcription was found among minor isoforms overexpressed in gHRD<sup>+</sup> relative to all such events in the genome. **B**, Nonfunctional minor transcripts overexpressed in gHRD<sup>+</sup> from ATM, BRIP1, MRE11, RAD51, and MRE11. **C**, RNA read profiles of aTU involving intron retention in representative genes in selected samples. Red, gHRD<sup>+</sup> samples; blue, gHRD<sup>-</sup> samples.

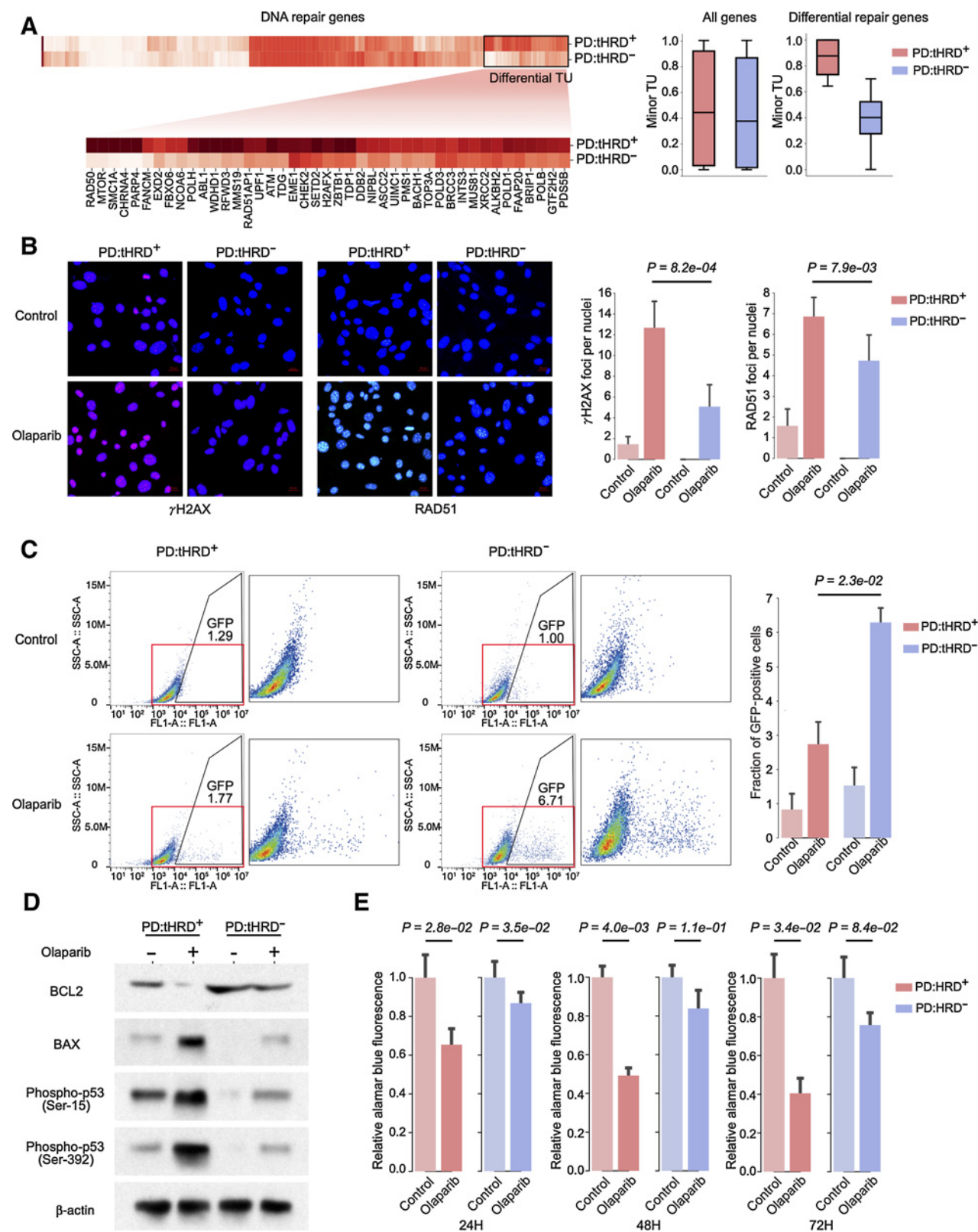
As expected, a majority (>87%) of BRCA1/2-mutant samples were gHRD<sup>+</sup> (Supplementary Fig. S1B, left). Importantly, a sizeable fraction (approximately 44%) of BRCA1/2-wildtype tumors also showed genomic instability at a stringent threshold, satisfying both metrics (Supplementary Fig. S1B, right), thereby implying HRD mechanisms that do not engage BRCA1/2 mutations. By using the transcriptomes of TCGA samples, we tested whether DNA repair genes were transcriptionally inactive in the gHRD<sup>+</sup> samples. However, the genes repressed in association with gHRD<sup>+</sup> were not enriched for molecular functions associated with DNA repair (Supplementary Fig. S1C).

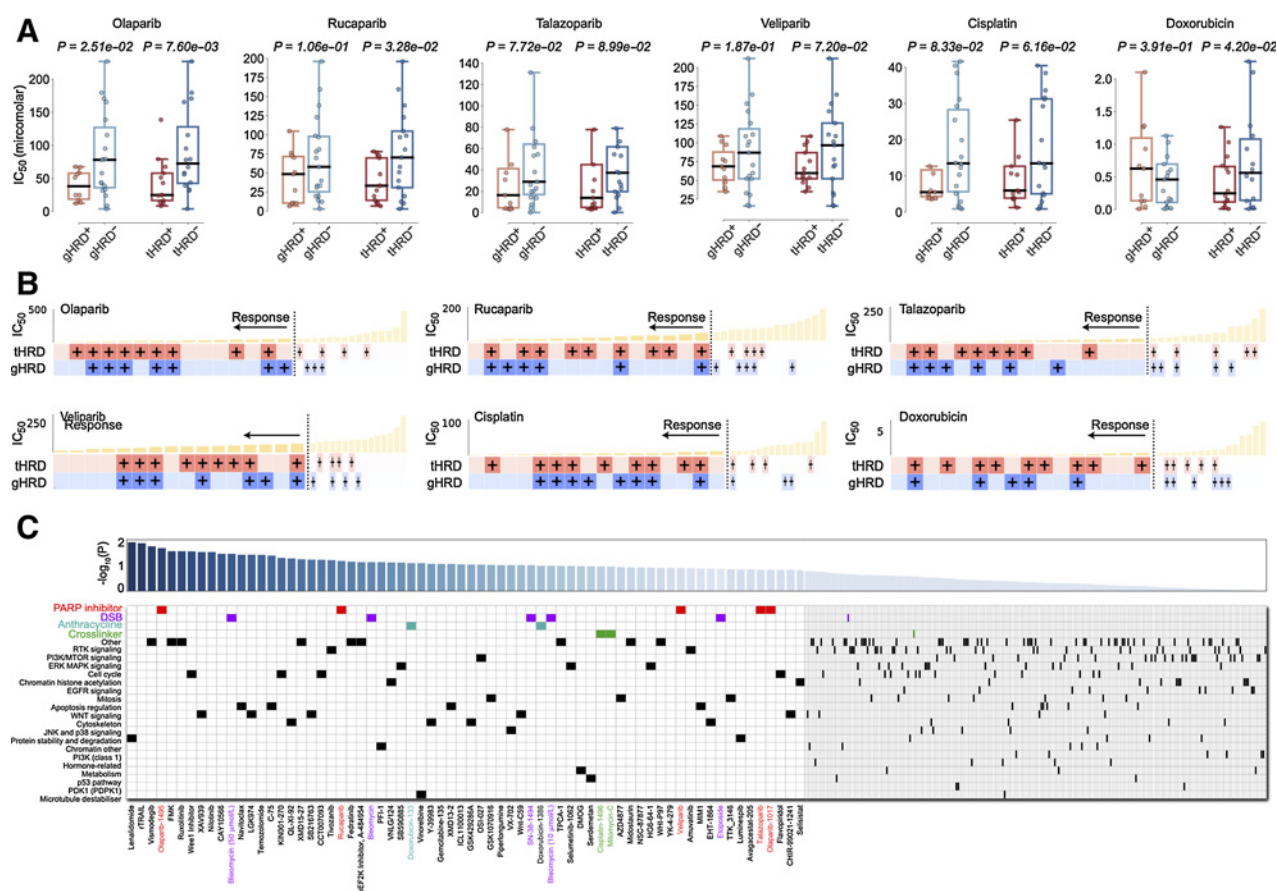
This prompted us to investigate differential RNA splicing between the gHRD<sup>+</sup> and gHRD<sup>-</sup> samples. Remarkably, our enrichment analysis indicated that genes involved in DNA repair tend to be spliced

differentially between the 2 groups (Supplementary Fig. S2A). However, splicing patterns themselves are not indicative of functional consequences. To further test whether splicing regulation is responsible for HRD, we examined the patterns of TU because minor or alternative isoforms can directly represent functional loss (32, 39). Specifically, we identified aTU as the relative overexpression of a minor transcript in the gHRD<sup>+</sup> samples (Fig. 1A and B). In total, we identified 2,521 such aTU genes (Fig. 1B) with a significant overrepresentation of DNA repair-related terms (Supplementary Fig. S2B), particularly DSB repair and DNA replication (red dots in Fig. 1C). The aTU patterns of representative DSB repair genes, including RAD51, are shown in Fig. 1D. Notably, this enrichment was not found for the opposite definition of aTU with minor isoform overexpression in gHRD<sup>-</sup> tumors (blue dots in Fig. 1C).



**Figure 3.** Functional validation of tHRD model based on genetic dependencies of clinical samples. **A**, Functional enrichment analysis of the genes predicted to be dependencies in TCGA tHRD<sup>+</sup> and tHRD<sup>-</sup> samples. Statistical significance of overrepresentation of the genes involved in various DNA repair pathways was plotted. The *q* values were calculated from the GSEApY enrichment test. **B**, Higher dependency scores in TCGA tHRD<sup>+</sup> and tHRD<sup>-</sup> samples (columns) were found for genes (rows) pertaining to DNA repair, recombination, and replication. Relevant functional categories are marked on the left side of the heatmap. Dependency scores for TCGA tumors were derived from our previous work (50). **C**, Examples of genes with higher dependency scores in the tHRD<sup>+</sup> and tHRD<sup>-</sup> tumors, respectively. Colored and gray curves represent the distribution of the tHRD<sup>+</sup> and tHRD<sup>-</sup> samples, respectively. Comparisons between two groups were performed using the Kolmogorov-Smirnov test. **D**, Genes (rows) whose minor isoform expression correlated with genetic dependencies on the PRKDC gene. Across TCGA samples (columns), we calculated the correlations between the PRKDC dependency score and the TU of minor isoforms of individual genes. Marked on the left are genes involved in the pathways of DNA repair, replication, and recombination.





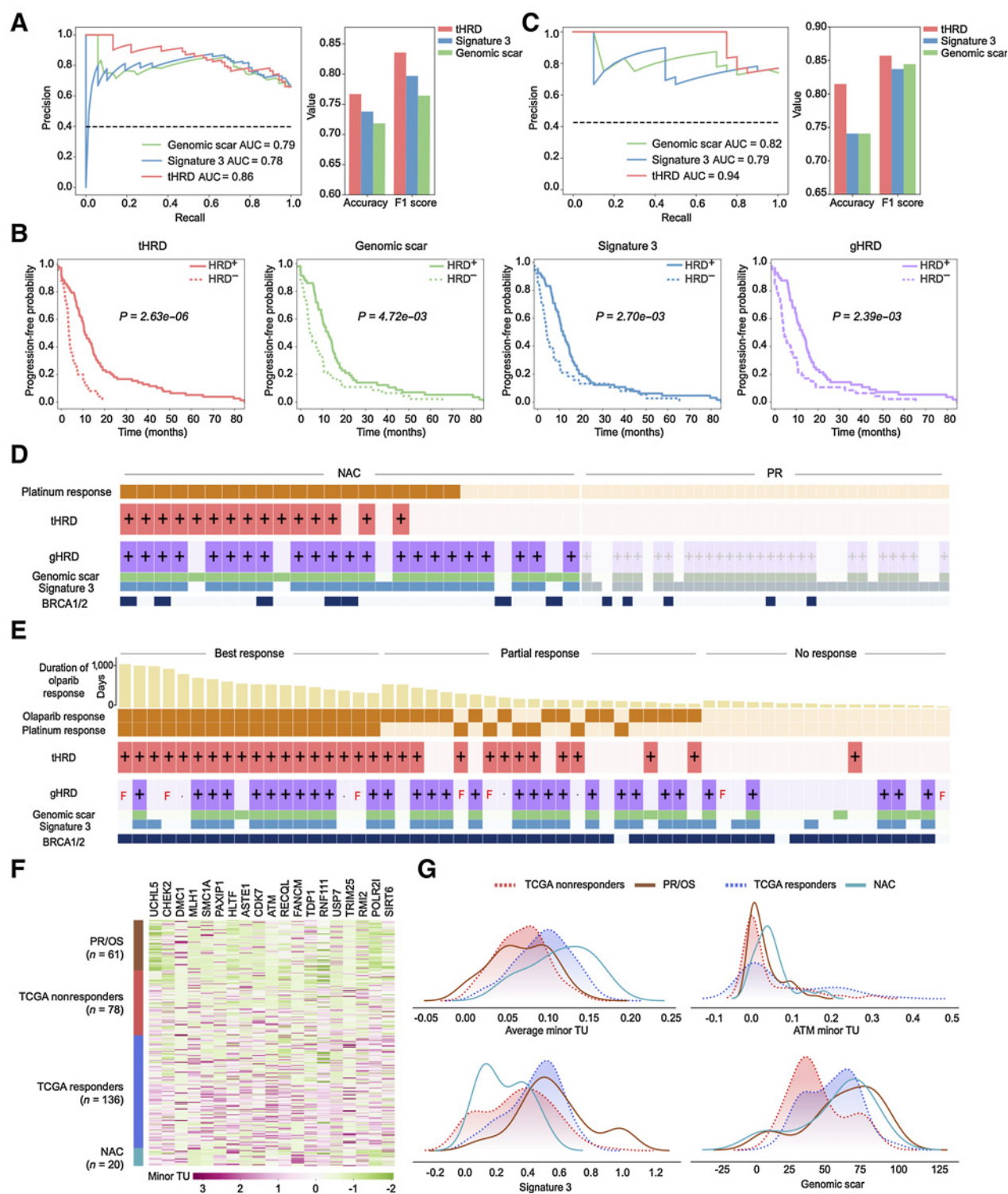
**Figure 5.** Responses to PARP inhibitors and genotoxic drugs explained by tHRD in cell lines. **A**, Comparison of  $IC_{50}$  between tHRD<sup>+</sup>/tHRD<sup>-</sup> versus gHRD<sup>+</sup>/gHRD<sup>-</sup> in breast cancer cell lines. The same graphs for the responses to bleomycin, etoposide, and SN-38 are provided in Supplementary Fig. S7. Response data were obtained from the GDSC database. We applied our tHRD detection model to the RNA-seq data of the cell lines to classify them as tHRD<sup>+</sup> and tHRD<sup>-</sup>. Comparisons between two groups were performed using the Mann-Whitney *U* test. **B**, Mapping of drug sensitivity and HRD prediction results in respective cell lines. Below the plot of  $IC_{50}$  values are the results of our tHRD model and signature 3-based gHRD assay for the corresponding cell lines. The plus signs denote HRD<sup>+</sup> predictions. gHRD was considered positive above the 75<sup>th</sup> percentile of signature 3.  $IC_{50}$  values lower than the median were regarded as positive responses to the given drug. **C**, Differences in drug sensitivity between the tHRD<sup>+</sup> and tHRD<sup>-</sup> breast cancer cell lines across drug classes. Differences are denoted as  $-\log_{10}P$  values at the top for drugs sorted in the order of the statistical significance. Below is the grouping of the drugs named at the bottom according to their mechanism of action shown on the left. PARP inhibitors, DSB-inducing agents, anthracyclines, and DNA cross-linkers are highlighted in colors. The *P* values are from the Mann-Whitney *U* test.

We examined whether the gHRD-associated aTU events give rise to functional loss. In fact, the frequency of retained introns and processed transcripts (i.e., noncoding transcripts that do not contain an open reading frame) by aTU was more than 3 times higher than expected from the genome-wide distribution (Fig. 2A). Nonfunctional minor

forms overexpressed in gHRD<sup>+</sup> from ATM, BRIP1, RAD51, and MRE11 are shown in Fig. 2B. For example, intron retention in BRIP1 results in the disruption of the BRCA1-interacting domain by a frameshift (Fig. 2B). We checked the RNA read profiles of aTU involving intron retention in BRIP1 and other genes in selected samples

**Figure 4.** Functional validation of tHRD model by HR assays in olaparib-treated PD cells. **A**, Comparison of the minor TU of DNA repair genes between PD tHRD<sup>+</sup> cells (PD:tHRD<sup>+</sup>) and tHRD<sup>-</sup> cells (PD:tHRD<sup>-</sup>). A cluster of genes showing minor form overexpression in PD:tHRD<sup>+</sup> was identified. Names of the genes are provided below the zoomed heatmap of differential TU. **B**, Immunofluorescence staining images for  $\gamma$ H2AX (red) and RAD51 (green) in PD:tHRD<sup>+</sup> and PD:tHRD<sup>-</sup> before and after olaparib treatment. Blue background indicates the fluorescent stain DAPI. Additional images are provided in Supplementary Fig. S6A. The graphs show the number of  $\gamma$ H2AX and Rad51 foci per cells, with the mean and SE obtained from four images. Comparisons between two groups were performed using the Student *t* test. **C**, Representative scatter plots showing the rate of HR repair based on the DR-GFP/I-Scel assay in PD:tHRD<sup>+</sup> and PD:tHRD<sup>-</sup> before and after olaparib treatment. GFP-positive cells in the marked and zoomed zones indicate a population of cells that underwent HR-mediated DSB repair of GFP reporter plasmids. The graph shows the fraction of the GFP-positive cells, with the mean and SE obtained from three replicate experiments after excluding the maximum and minimum. Comparisons between two groups were performed using the Student *t* test. Replicate plots are provided in Supplementary Fig. S6B. **D**, Western blot analysis of PD:tHRD<sup>+</sup> and PD:tHRD<sup>-</sup> before and after olaparib treatment for BCL2 (antiapoptotic protein), BAX (proapoptotic protein), P53 phosphorylation at Ser-15 (senescence marker), and P53 phosphorylation at Ser-392 (apoptosis marker). **E**, Cell viability measured by the alamarBlue assay in PD:tHRD<sup>+</sup> and PD:tHRD<sup>-</sup> before and after olaparib treatment. Fluorescence signals were obtained at 24, 48, and 72 hours after the treatment and normalized by the pretreatment (control) measurements. Mean and SE were obtained from three experiments. Comparisons between two groups were performed using the Student *t* test. H, hours.





**Figure 6.** Therapeutic responses to olaparib and platinum regimens explained by tHRD. **A** and **B**, HRD-based prediction of platinum sensitivity of patients with ovarian cancer in TCGA ( $n = 103$ ). **A**, Precision versus recall (left), and accuracy and F1 at the best accuracy (right), for tHRD, genomic scar, and signature 3. **B**, PFS for patients partitioned by the tHRD- and gHRD-based classification. PFI comparisons between two groups were performed using the log-rank test. **C**, Precision versus recall (left), and accuracy and F1 at the best accuracy (right), for tHRD, genomic scar, and signature 3 in NAC ( $n = 27$ ). **D**, Mapping of platinum responses and HRD prediction results in respective NAC ( $n = 27$ ) and PR ( $n = 36$ ) samples. For tHRD and signature 3, the thresholds at the best accuracy were obtained from NAC and applied to PR. For genomic scar, the previously established threshold for clinical use (65) was employed to reduce false positives in PR. (Continued on the following page.)

(red for gHRD<sup>+</sup> samples and blue for gHRD<sup>-</sup> samples in Fig. 2C). The aTU patterns were observed specifically in gHRD<sup>+</sup> tumors but not in their matched normal samples (Supplementary Fig. S3).

### Development and functional validation of the tHRD detection model

We selected 35 DNA repair genes that showed aTU in TCGA breast cancer and developed a tHRD prediction model based on the TU level of their transcripts (Supplementary Fig. S4A; Supplementary Table S1). The performance of the aTU-based classifier in predicting gHRD indicates only the degree of agreement between the tHRD and gHRD methods. Some gHRD<sup>+</sup> samples may be predicted as tHRD<sup>-</sup> when they present artefactual genomic signatures as described in the Introduction. Also, some gHRD<sup>-</sup> samples may be predicted to be tHRD<sup>+</sup> when they share similar TU patterns with genuine gHRD<sup>+</sup> samples; these should not be considered as false positives. Hence, the tHRD results should be validated by functional or clinical measures independently of gHRD status. For example, HRD<sup>+</sup> tumors are expected to show poor prognosis and high mutation burden. In this regard, the tHRD approach was as successful as the gHRD method (Supplementary Fig. S4B).

To perform more detailed functional tests of the tHRD results in a gHRD-independent manner, we first employed the concept of genetic dependencies. Substantial effort has been put into identifying genes essential for cancer cell proliferation and survival based on systematic loss-of-function screens (40–49). Synthetic lethality arises when inactivation of one gene (e.g., by BRCA1/2 mutations) increases dependencies on a related gene (e.g., PARP1) so that the simultaneous perturbation of the two genes results in cell death (7). In this respect, the samples predicted to be HRD<sup>+</sup> should be dependent on genes with DNA repair activity.

We first tested this concept by using cell lines for which both transcriptome and dependency data were available. Cell lines with higher tHRD scores represented higher average dependencies on DNA repair genes (Supplementary Fig. S5). For validation in clinical samples, we performed dependency prediction (50) for TCGA tHRD<sup>+</sup> and tHRD<sup>-</sup> samples, and then examined enrichment of the predicted dependencies for DNA repair genes. The dependencies of the tHRD<sup>+</sup> tumors, but not the tHRD<sup>-</sup> tumors, showed a significant overrepresentation of genes involved in DNA repair pathways (red versus blue bars in Fig. 3A). Indeed, higher dependency scores were observed in tHRD<sup>+</sup> than in tHRD<sup>-</sup> for many DNA repair genes, with the largest difference for PRKDC (Fig. 3B and C). As the main component of the nonhomologous end joining pathway, DNA-PKcs (encoded by PRKDC) is recruited to DSBs for DNA repair when HR is unavailable or defective (51). Further studies established synthetic lethality between the HR pathway and DNA-PKcs (52–55). We discovered many genes involved in DNA repair and related pathways showing positive correlations between minor TU and PRKDC dependency (Fig. 3D).

We next sought to validate our tHRD prediction by functional HR assays in BRCA1/2-wildtype patient-derived (PD) cells. One tHRD<sup>+</sup> sample (PD:tHRD<sup>+</sup>) and one tHRD<sup>-</sup> sample (PD:tHRD<sup>-</sup>) were selected by applying the tHRD model to the transcriptomes of our PD xenografts of breast cancer (56). PD:tHRD<sup>+</sup> overexpressed the

minor isoforms of key repair genes, including RAD50, ATM, H2AX, and BRIP1 (Fig. 4A). In our first assay, DSB repair processes were visualized by  $\gamma$ H2AX and RAD51 staining before and after olaparib treatment. Whereas olaparib treatment induced DSBs overall, a significantly larger number of  $\gamma$ H2AX and RAD51 foci were observed in PD:tHRD<sup>+</sup> than in PD:tHRD<sup>-</sup> (Fig. 4B; Supplementary Fig. S6A). For a more direct assessment of HR-mediated DSB repair activity, we performed the DR-GFP/I-SceI HR assay (57) in which GFP signals represent the full repair of introduced DSBs. As a result, PD:tHRD<sup>+</sup> showed a 2- to 3-fold lower fraction of GFP-positive cells than did PD:tHRD<sup>-</sup> (Fig. 4C; Supplementary Fig. S6B). These results collectively suggest that the accumulation of  $\gamma$ H2AX/RAD51 foci in the PD:tHRD<sup>+</sup> cells results from delayed DSB resolution due to impaired HR functionality, as demonstrated previously (58–61).

To compare the cellular consequences of the differential rates of HR repair, we examined the phosphorylation of P53, the key player of DDR. The phosphorylation of P53 at Ser-392 and at Ser-15 gives rise to apoptosis and cellular senescence, respectively (62, 63). According to the Western blots, olaparib treatment strongly induced P53 phosphorylation, particularly at Ser-392 in PD:tHRD<sup>+</sup> but not in PD:tHRD<sup>-</sup> (Fig. 4D). Accordingly, PD:tHRD<sup>+</sup> presented an increase of the proapoptotic protein, BAX, and a decrease of the antiapoptotic protein, BCL2, in response to olaparib (Fig. 4D). These data suggest that the unresolved DSB foci of PD:tHRD<sup>+</sup> (Fig. 4B and C) eventually lead to apoptosis of the cells. Indeed, our cell viability assays showed a significant reduction in cell proliferation when the PD:tHRD<sup>+</sup> cells were challenged by olaparib (Fig. 4E).

### tHRD predicts drug responses with high accuracy in cell lines

To evaluate our tHRD model in predicting susceptibility to PARP inhibition or genotoxic stress, we examined the response of breast cancer cell lines to olaparib, rucaparib, talazoparib, veliparib, cisplatin, doxorubicin, bleomycin, etoposide, and SN-38 using the Genomics of Drug Sensitivity in Cancer (GDSC) database (64). We applied our tHRD detection model to the RNA-seq data of the cell lines to classify them as tHRD<sup>+</sup> and tHRD<sup>-</sup> (Supplementary Table S2). In terms of the IC<sub>50</sub>, significant differences were found in the sensitivity of the tHRD<sup>+</sup> and tHRD<sup>-</sup> cell lines for all the PARP inhibitors and DNA damaging agents (Fig. 5A; Supplementary Fig. S7A and S7B).

Remarkably, the IC<sub>50</sub> differences were more significant between the tHRD<sup>+</sup> and tHRD<sup>-</sup> groups than between the gHRD<sup>+</sup> and gHRD<sup>-</sup> groups (Fig. 5A) for almost all drugs, indicating that tHRD better predicts drug sensitivity than gHRD. For instance, the olaparib and rucaparib responses scored  $P = 0.008$  (tHRD) versus  $P = 0.025$  (gHRD) and  $P = 0.033$  (tHRD) versus  $P = 0.106$  (gHRD), respectively. We then examined the tHRD and gHRD status of individual samples. Notably, there were multiple cases in which the response was predicted only by tHRD but not by gHRD (i.e., tHRD<sup>+</sup>/gHRD<sup>-</sup> samples) in the responsive group (IC<sub>50</sub> values lower than the median; Fig. 5B). Lowering the threshold of gHRD (from the 75<sup>th</sup> to 50<sup>th</sup> percentile) led to many cases in which only gHRD made a positive prediction (i.e., gHRD<sup>+</sup>/tHRD<sup>-</sup> samples) in the no response group (IC<sub>50</sub> values greater than the median; Supplementary Fig. S8A). At this threshold, however, gHRD did not make a meaningful distinction between the

(Continued.) The plus signs for gHRD denote the cases in which both genomic scar and signature 3 were positive. Genetic screening results indicating the presence of germline mutations in BRCA1/2 or CHEK2 are shown at the bottom. **E**, Mapping of olaparib/platinum responses and HRD prediction results at the best accuracy in respective OM ( $n = 24$ ) and OS ( $n = 33$ ) samples. Duration of olaparib responses is shown at the top. Platinum responses are marked as positive for the patients with OM and negative for the patients with OS. The plus signs for gHRD denote the cases in which both genomic scar and signature 3 were positive. **F**, genome sequencing failure. Genetic screening results indicating the presence of germline mutations in BRCA1/2 or other genes (ATM, CHEK1, and CDK12) are shown at the bottom. **F**, Comparison of the minor TU of selected genes (columns) across different samples (rows) from the OS/PR, TCGA, and NAC cohorts. Genes with the most significant differences between TCGA responders and nonresponders were selected. **G**, Comparison of the distribution of tHRD and gHRD metrics among the OS/PR, TCGA, and NAC groups of samples. For the average minor TU, the TU values of the minor isoform of the selected genes in **F** were averaged for each sample.

positive and negative samples (see the *P* values above the boxplots of Supplementary Fig. S8B).

These data suggest that gHRD is prone to false positive signals attributed to biological or technical artefacts (29). This feature of gHRD explains low levels of precision at a common cut-off of the median with a considerable number of the gHRD<sup>+</sup> samples that do not actually respond to the drugs (Supplementary Fig. S8A). A more stringent threshold improves precision but sacrifices recall (also known as sensitivity), as demonstrated in Fig. 5B. In this case, the tHRD approach is more sensitive than the gHRD method and can identify tumors that do not harbor detectable genomic signatures but that respond to these drugs.

Additionally, we tested the drug specificity of the responses as reported previously (14) by comparing IC<sub>50</sub> among different classes of molecules grouped according to their mechanism of action. Specific responses were observed for PARP inhibitors (olaparib, rucaparib, veliparib, and talazoparib; red in Fig. 5C), DSB-inducing agents (bleomycin, SN-38, and etoposide; violet in Fig. 5C), and anthracyclines or DNA cross-linkers (doxorubicin, cisplatin, and mitomycin C; green in Fig. 5C).

The number of ovarian cancer cell lines with IC<sub>50</sub> in the GDSC database was limited. Hence, we merged data based on the area under the drug response curve from different databases. As a result, gHRD did not show an expected pattern (i.e., lower AUC value for HRD<sup>+</sup>) for all the examined drugs. In contrast, tHRD consistently showed the expected pattern although statistical significance was not achieved at *P* < 0.1 except for olaparib and carboplatin (Supplementary Fig. S7C).

### tHRD predicts drug responses with high accuracy in clinical samples

We next sought to assess the utility of tHRD in clinical settings. To first develop a tHRD model for ovarian cancer, the procedures we applied to TCGA breast tumors were repeated for TCGA ovarian samples; briefly, we partitioned the patient samples into gHRD<sup>+</sup> and gHRD<sup>-</sup> groups, identified aTU cases with minor isoform overexpression in gHRD<sup>+</sup>, and developed an aTU-based prediction model with 22 DNA repair genes.

By using the reannotation of clinical outcomes of TCGA cases (*n* = 103; ref. 33), we evaluated the performance of our tHRD model in predicting platinum sensitivity (Supplementary Table S3). In comparison with gHRD, the tHRD prediction exhibited an improvement in terms of the precision and recall metrics as well as the accuracy and F1 score (Fig. 6A). Also, PFS was substantially better explained by tHRD (*P* = 2.6 × 10<sup>-6</sup>) than genomic scar (*P* = 4.7 × 10<sup>-3</sup>), signature3 (*P* = 2.7 × 10<sup>-3</sup>), and their combination (*P* = 2.4 × 10<sup>-3</sup>; Fig. 6B). We further tested the model using samples from our NAC cohort (*n* = 27) and performed validation by adding independent samples with PR (*n* = 36; Supplementary Table S4). The tHRD model outperformed gHRD in the NAC cohort (Fig. 6C) and made no false positives in the PR cohort (Fig. 6D; Supplementary Fig. S9A). In contrast, gHRD made a substantial number of false positive predictions for the PR samples (Fig. 6D), failing to achieve a clinically applicable level of accuracy for the combined dataset (Supplementary Fig. S9A). BRCA1/2 testing was not successful at all in predicting platinum responses (Fig. 6D).

We then tested whether the tHRD model can predict responses to PARP inhibitors as well by using our data from OM and OS treatment (Supplementary Table S5). The OM group (*n* = 24) and OS group (*n* = 33) data represented therapeutic responses of platinum-sensitive and platinum-resistant patients, respectively, to olaparib as second-line therapy. Whereas tHRD outperformed gHRD in predicting olaparib sensitivity overall (Fig. 6E; Supplementary Fig. S9B), we focused on the

cases that were resistant to both olaparib and platinum (*n* = 17). Notably, both genomic scar and signature 3 were positive for 7 cases in this highly resistant group in which only one tHRD<sup>+</sup> case was found (Fig. 6E). BRCA1/2 screening also resulted in many false positives (Fig. 6E). Demonstrating the coupling of precision to sensitivity, the gHRD methods failed to detect a few of the best responders that were sensitive to both olaparib and platinum (Fig. 6E). Among patients with partial response, sensitive to either olaparib or platinum, tHRD was positive for those with a relatively longer duration of olaparib response (Fig. 6E), thereby contributing to the clear segregation of survival curves (Supplementary Fig. S10).

To test whether the platinum resistance of the PR and OS cases can be explained by TU patterns, we partitioned TCGA samples into platinum responders and non-responders and identified key repair genes whose minor isoforms were underexpressed in the nonresponders. The TU of these minor isoforms in the PR/OS samples was as low as the nonresponders (Fig. 6F). The distribution of the average TU of these isoforms and that of the TU of the ATM minor isoform were similar between the PR/OS and TCGA nonresponder populations (top panel of Fig. 6G). In sharp contrast, both signature 3 and genomic scar of the PR/OS samples were distributed similarly to those of TCGA responders (bottom panel of Fig. 6G).

Together with our findings about breast cancer (Supplementary Fig. S8), these findings about ovarian cancer underscore the low precision of gHRD due to confounding factors that can be either biological or technical (29). The false positives in the PR cohort may be technical artefacts, given that most of the patients are BRCA1/2-wildtype (Fig. 6D). In contrast, our OS data highlight the influence of biological factors, considering that the olaparib users were selected based on their BRCA1/2 mutations. In this respect, the nonresponders of this group are likely to involve initially imprinted genomic lesions that remain detectable after HR restoration. Given that acquired platinum resistance is common in ovarian cancer (22–25), this raises a problem when attempting to select patient cases that will benefit from PARP inhibitors as subsequent therapy based on the gHRD signatures or BRCA1/2 mutations. In contrast, the tHRD model predicted durable responders sensitively while maintaining high precision.

The tHRD data from some of the OM and OS samples were derived from FFPE tissues. It is known that nucleic acids are not preserved very well in FFPE tissues. Indeed, we failed DNA sequencing for some of the FFPE samples, resulting in missing data for gHRD. In contrast, our results suggest that the RNA-based tHRD assay may be applied properly to these most common sources of archived tissue material.

## Discussion

PARP inhibition is a promising treatment option targeted to cancers in which DSB repair has been impaired by mechanisms such as BRCA1/2 inactivation. However, BRCA1/2 mutations comprise only a small fraction of HRD cases. Our data reinforce the notion that many HRD cases do not engage BRCA1/2 alterations. Because of the diversity of potential mechanisms, however, there has been no single unified mechanism-based assay with greater utility than the genetic screening of BRCA1/2 or HR repair gene mutations. The gHRD assays, which scan the consequences of genomic instability caused by HR defects, lack functional flexibility and suffer a high rate of false positives.

In this work, we focused on a less explored class of transcriptional aberration and found that these alterations in DNA repair-related genes prevail in breast and ovarian cancer without regard to BRCA1/2 mutations. On the basis of this widespread mechanism, we developed a

novel approach that has the capacity to substitute for not only the genetic screening but also the gHRD assays. The tHRD method appears to reflect functional HR status that changes dynamically during cancer evolution, thereby minimizing biological confounding factors, in contrast to the static genome-based tests.

What remains to be tested is whether this HRD mechanism is less reversible than other mechanisms, especially under therapeutic interventions. Investigation of underlying causal factors will be helpful in this respect. Of note, these transcriptional changes involve multiple genes with sample-to-sample variations, ruling out the possibility that a single splicing factor acting in *trans* governs this somewhat complicated phenomenon; thus, genetic or epigenetic regulatory changes influencing respective genes in *cis* are likely more responsible. We employed a machine learning-based model to address this variability. In this regard, the reversion of the HRD driven in this fashion is less likely to occur than that involving single-gene coding mutations.

The utility of this method remains to be validated further beyond the multiple different contexts and settings tested in this work. Our method is expected to offer high adaptability as machine learning models can be optimized by adjusting various features and parameters. Especially, the tHRD model can be fitted directly to particular therapeutic responses instead of the universal genomic markers, as long as sufficient training data can be provided. For example, different repair genes can be used as features with variable parameters in predicting responses to different drugs in various clinical settings. In this work, we developed respective models for breast cancer and ovarian cancer. The tHRD-based diagnostic tests are expected to broaden the clinical utility of PARP inhibitors in various cancers.

### Authors' Disclosures

H.G. Kang reports a patent for Method for determining sensitivity to PARP inhibitor or genotoxic drugs based on non-functional transcripts pending. E-H. Cho reports grants from Ministry of Science and ICT during the conduct of the study; in

addition, E. Cho has a patent for Method for determining sensitivity to PARP inhibitor or genotoxic drugs based on non-functional transcripts pending. J-Y. Lee reports grants from Ministry of Trade, Industry and Energy; grants and personal fees from AstraZeneca, Takeda; and grants and personal fees from MSD during the conduct of the study. J.K. Choi reports a patent for Method for determining sensitivity to PARP inhibitor or genotoxic drugs based on non-functional transcripts pending. No disclosures were reported by the other authors.

### Authors' Contributions

**H.G. Kang:** Conceptualization, data curation, formal analysis, investigation, visualization, methodology, writing—original draft. **H. Hwangbo:** Conceptualization, formal analysis, investigation. **M.J. Kim:** Formal analysis. **S. Kim:** Formal analysis. **E.J. Lee:** Formal analysis. **M.J. Park:** Formal analysis. **J.-W. Kim:** Resources, project administration. **B.-G. Kim:** Resources, project administration. **E.-H. Cho:** Resources, data curation, supervision, funding acquisition, project administration. **S. Chang:** Formal analysis, supervision, validation, investigation, project administration. **J.-Y. Lee:** Conceptualization, resources, data curation, formal analysis, supervision, project administration. **J.K. Choi:** Supervision, funding acquisition.

### Acknowledgments

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF-2017M3A9A7050612) and by the Innovation Growth Engine for Planning and Demonstration of the Commercialization Promotion Agency For R&D Outcomes (NTIS-1711121252) funded by the Korean government (MSIT).

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

### Note

Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Received June 29, 2021; revised September 15, 2021; accepted October 25, 2021; published first October 28, 2021.

### References

- 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.
- Gao D, Herman JG, Guo M. The clinical value of aberrant epigenetic changes of DNA damage repair genes in human cancer. *Oncotarget* 2016;7:37331–46.
- Riaz N, Blecua P, Lim RS, Shen R, Higginson DS, Weinhold N, et al. Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. *Nat Commun* 2017;8:1–7.
- Farmer H, McCabe H, Lord CJ, Tutt AHJ, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005;434:917–21.
- Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 2005;434:913–7.
- Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 2009;361:123–34.
- O'Neil NJ, Bailey ML, Hieter P. Synthetic lethality and cancer. *Nat Rev Genet* 2017;18:613–23.
- Lord CJ, Ashworth A. BRCAness revisited. *Nat Rev Cancer* 2016;16:110–20.
- Lord CJ, Ashworth A. PARP inhibitors: synthetic lethality in the clinic. *Science* 2017;355:1152–8.
- 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;13:517–25.
- 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.
- 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.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 2013;3:246–59.
- Gulhan DC, Lee JJK, Melloni GEM, Cortés-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. *Nat Genet* 2019;51:912–9.
- Staaf J, Glodzik D, Bosch A, Vallon-Christersson J, Reuterswärd C, Häkkinen J, et al. Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. *Nat Med* 2019;25:1526–33.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–21.
- Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer* 2012;107:1776–82.
- Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* 2012;72:5454–62.
- Birnbak NJ, Wang ZC, Kim J-Y, Eklund AC, Li Q, Tian R, et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discov* 2012;2:366–75.
- Lord CJ, Ashworth A. Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nat Med* 2013;19:1381–8.
- Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* 2015;521:489–94.

22. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C, et al. Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* 2008;451:1116–20.
23. Edwards SL, Brough R, Lord CJ, Natrajan R, Vatcheva R, Levine DA, et al. Resistance to therapy caused by intragenic deletion in BRCA2. *Nature* 2008;451:1111–5.
24. Norquist B, Wurz KA, Pennil CC, Garcia R, Gross J, Sakai W, et al. Secondary somatic mutations restoring BRCA1/2 predict chemotherapy resistance in hereditary ovarian carcinomas. *J Clin Oncol* 2011;29:3008–15.
25. Swisher EM, Sakai W, Karlan BY, Wurz K, Urban N, Taniguchi T. Secondary BRCA1 mutations in BRCA1-mutated ovarian carcinomas with platinum resistance. *Cancer Res* 2008;68:2581–6.
26. Chaudhuri AR, Callen E, Ding X, Gogola E, Duarte AA, Lee JE, et al. Replication fork stability confers chemoresistance in BRCA-deficient cells. *Nature* 2016;535:382–7.
27. Xu G, Ross Chapman J, Brandsma I, Yuan J, Mistrik M, Bouwman P, et al. REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature* 2015;521:541–4.
28. Maura F, Degasperis A, Nadeu F, Leongamornlert D, Davies H, Moore L, et al. A practical guide for mutational signature analysis in hematological malignancies. *Nat Commun* 2019;10:2969.
29. Watkins JA, Irshad S, Grigoriadis A, Tutt ANJ. Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers. *Breast Cancer Res* 2014;16:211.
30. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;33:290–5.
31. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Research* 2020;9:1–19.
32. Rodriguez JM, Rodriguez-Rivas J, Di Domenico T, Vázquez J, Valencia A, Tress ML. APPRIS 2017: principal isoforms for multiple gene sets. *Nucleic Acids Res* 2018;46:D213–7.
33. Villalobos VM, Wang YC, Sikic BI. Reannotation and analysis of clinical and chemotherapy outcomes in the ovarian data set from The Cancer Genome Atlas. *JCO Clin Cancer Informatics* 2018;2:1–16.
34. Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, et al. Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *N Engl J Med* 2012;366:1382–92.
35. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–24.
36. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* 2016;17:31.
37. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* 2015;26:64–70.
38. 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;4:8–11.
39. Rodriguez JM, Maietta P, Ezkurdia I, Pietrelli A, Wesselink JJ, Lopez G, et al. APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res* 2013;41:D110–7.
40. Hart T, Chandrashekar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 2015;163:1515–26.
41. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, et al. Identification and characterization of essential genes in the human genome. *Science* 2015;350:1096–101.
42. Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, et al. Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. *Cell* 2016;164:293–309.
43. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. *Cell* 2017;170:564–76.
44. McDonald ER, de Weck A, Schlabach MR, Billy E, Mavrakis KJ, Hoffman GR, et al. Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. *Cell* 2017;170:577–92.
45. Wang T, Yu H, Hughes NW, Liu B, Kendirli A, Klein K, et al. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* 2017;168:890–903.
46. Patel SJ, Sanjana NE, Kishton RJ, Eidizadeh A, Vodnala SK, Cam M, et al. Identification of essential genes for cancer immunotherapy. *Nature* 2017;548:537–42.
47. McFarland JM, Ho ZV, Kugener G, Dempster JM, Montgomery PG, Bryan JG, et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat Commun* 2018;9:1–13.
48. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* 2017;49:1779–84.
49. Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* 2019;568:511–6.
50. Jang K, Park MJ, Park JS, Hwangbo H, Sung MK, Kim S, et al. Computational inference of cancer-specific vulnerabilities in clinical samples. *Genome Biol* 2020;21:155.
51. Spagnolo L, Rivera-Calzada A, Pearl LH, Llorca O. Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol Cell* 2006;22:511–9.
52. Gurley KE, Kemp CJ. Synthetic lethality between mutation in Atm and DNA-PK(cs) during murine embryogenesis. *Curr Biol* 2001;11:191–4.
53. Riabinska A, Daheim M, Herter-Sprie GS, Winkler J, Fritz C, Hallek M, et al. Therapeutic targeting of a robust non-oncogene addiction to PRKDC in ATM-defective tumors. *Sci Transl Med* 2013;5:189ra78.
54. Dietlein F, Thelen L, Reinhardt HC. Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. *Trends Genet* 2014;30:326–39.
55. Dietlein F, Thelen L, Jokic M, Jachimowicz RD, Ivan L, Knittel G, et al. A functional cancer genomics screen identifies a druggable synthetic lethal interaction between MSH3 and PRKDC. *Cancer Discov* 2014;4:592–605.
56. Jung J, Jang K, Ju JM, Lee E, Lee JW, Kim HJ, et al. Novel cancer gene variants and gene fusions of triple-negative breast cancers (TNBCs) reveal their molecular diversity conserved in the patient-derived xenograft (PDX) model. *Cancer Lett* 2018;428:127–38.
57. Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev* 1999;13:2633–8.
58. Panier S, Boulton SJ. Double-strand break repair: 53BP1 comes into focus. *Nat Rev Mol Cell Biol* 2014;15:7–18.
59. Chapman JR, Barral P, Vannier JB, Borel V, Steger M, Tomas-Loba A, et al. RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Mol Cell* 2013;49:858–71.
60. Ayoub N, Rajendra E, Su X, Jeyasekharan AD, Mahen R, Venkitaraman AR. The carboxyl terminus of Brca2 links the disassembly of Rad51 complexes to mitotic entry. *Curr Biol* 2009;19:1075–85.
61. Shimizu N, Akagawa R, Takeda S, Sasanuma H. The MRE11 nuclease promotes homologous recombination not only in DNA double-strand break resection but also in post-resection in human TK6 cells. *Genome Instab Dis* 2020;1:184–96.
62. Achison M, Hupp TR. Hypoxia attenuates the p53 response to cellular damage. *Oncogene* 2003;22:3431–40.
63. Lindström MS, Wiman KG. Myc and E2F1 induce p53 through p14ARF-independent mechanisms in human fibroblasts. *Oncogene* 2003;22:4993–5005.
64. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;41:955–61.
65. Melinda LT, Kirsten MT, Julia R, Bryan H, Gordon BM, Kristin CJ, 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.