

# RNA-Based Classification of Homologous Recombination Deficiency in Racially Diverse Patients with Breast Cancer

Andrea Walens<sup>1,2</sup>, Sarah C. Van Alsten<sup>2</sup>, Linnea T. Olsson<sup>2</sup>, Markia A. Smith<sup>3</sup>, Alex Lockhart<sup>4</sup>, Xiaohua Gao<sup>1,2</sup>, Alina M. Hamilton<sup>3</sup>, Erin L. Kirk<sup>2</sup>, Michael I. Love<sup>4,5</sup>, Gaorav P. Gupta<sup>1</sup>, Charles M. Perou<sup>1,5</sup>, Cyrus Vaziri<sup>3</sup>, Katherine A. Hoadley<sup>1,5</sup>, and Melissa A. Troester<sup>1,2,3</sup>



## ABSTRACT

**Background:** Aberrant expression of DNA repair pathways such as homologous recombination (HR) can lead to DNA repair imbalance, genomic instability, and altered chemotherapy response. DNA repair imbalance may predict prognosis, but variation in DNA repair in diverse cohorts of breast cancer patients is understudied.

**Methods:** To identify RNA-based patterns of DNA repair expression, we performed unsupervised clustering on 51 DNA repair-related genes in the Cancer Genome Atlas Breast Cancer [TCGA BRCA ( $n = 1,094$ )] and Carolina Breast Cancer Study [CBCS ( $n = 1,461$ )]. Using published DNA-based HR deficiency (HRD) scores (high-HRD  $\geq 42$ ) from TCGA, we trained an RNA-based supervised classifier. Unsupervised and supervised HRD classifiers were evaluated in association with demographics, tumor characteristics, and clinical outcomes.

**Results :** Unsupervised clustering on DNA repair genes identified four clusters of breast tumors, with one group having high expression of HR genes. Approximately 39.7% of CBCS and 29.3% of TCGA breast tumors had this unsupervised high-HRD (U-HRD) profile. A supervised HRD classifier (S-HRD) trained on TCGA had 84% sensitivity and 73% specificity to detect HRD-high samples. Both U-HRD and S-HRD tumors in CBCS had higher frequency of TP53 mutant-like status (45% and 41% enrichment) and basal-like subtype (63% and 58% enrichment). S-HRD high was more common among black patients. Among chemotherapy-treated participants, recurrence was associated with S-HRD high (HR: 2.38, 95% confidence interval = 1.50–3.78).

**Conclusions:** HRD is associated with poor prognosis and enriched in the tumors of black women.

**Impact:** RNA-level indicators of HRD are predictive of breast cancer outcomes in diverse populations.

## Introduction

Aberrant expression of certain DNA repair pathways, especially homologous recombination (HR), can lead to DNA repair imbalance, genomic instability, and altered chemotherapy response (1–5). Specifically, it has been previously shown that tumors with HR deficiency (HRD) are more sensitive to platinum-containing and DNA damage agents, and treatment of HRD<sup>+</sup> patients show increased response rate and prolonged survival (6–8). Much of the focus in targeting HRD has been on estrogen receptor (ER) negative or triple-negative breast cancer (TNBC; refs. 9, 10). For ER<sup>+</sup> breast cancer, St. Gallen's

recommends chemotherapy only for high grade and stage (11) or for tumors with high genomic risk of recurrence (12, 13). Most currently available genomic tests are based on overall proliferation and not specific DNA-repair pathways (14), and therefore, it would be valuable to identify DNA repair pathways that represent molecular vulnerabilities to chemotherapy (15) in estrogen receptor-positive (ER<sup>+</sup>) breast cancer (16).

The role of DNA repair is also poorly understood in diverse populations; however, several studies have shown that black breast cancer patients tend to receive chemotherapy at higher rates than non-black patients (17) and one recent study has demonstrated differences in DNA repair pathway expression by race (18). We interpret race herein as a social construct that acts on multiple levels, from cells to society. Different prevalence of breast cancer subtypes may result from unequal exposures to environmental carcinogens, differential access to treatment, screening, and preventative care, disparities in economic access, ancestry, and other factors associated with racism and social determinants of health.

Second, most studies of candidate biomarkers for HRD have focused on using DNA-based data. Using DNA repair pathway aberrations measured by whole-exome or genome sequencing (WES/WGS), studies have defined presence/absence of mutational signatures (19) that integrate across genes. This is an attractive approach for identifying samples with molecular vulnerability because it does not rely on mutations in specific genes, instead identifying patterns of alterations associated with specific DNA repair processes or exposures. However, in situations where WES/WGS data are unavailable or of low-quality, alternate pathway-based approaches of monitoring the HR pathway may supplement findings based on DNA evidence.

RNA-based expression of DNA repair genes is a convenient approach for tracking DNA repair pathway activity. Other RNA-based classifiers, such as those for TP53 mutant-like status (20, 21),

<sup>1</sup>Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina. <sup>2</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina. <sup>3</sup>Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina, Chapel Hill, North Carolina. <sup>4</sup>Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina. <sup>5</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

A. Walens and S.C. Van Alsten contributed equally as the co-first authors of this article.

**Corresponding Author:** Melissa A. Troester, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, NC, 27599. Phone: 919-966-7409; E-mail: troester@unc.edu

Cancer Epidemiol Biomarkers Prev 2022;31:2136–47

doi: 10.1158/1055-9965.EPI-22-0590

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

©2022 The Authors; Published by the American Association for Cancer Research

exemplify that RNA-level data can indicate DNA alterations and complement somatic mutation data. For instance, RNA-based TP53 classifiers accurately identify TP53 DNA mutations and predict survival with hazard ratios that are similar to those from DNA-based assays (20). We hypothesized that relative expression levels of core mediators of error-free and error-prone genome maintenance pathways could be used to build RNA expression-based classifiers for DNA repair pathways. We aimed to develop a gene expression classifier for HRD and apply it across multiple human and model system breast cancer datasets, including the Carolina Breast Cancer Study (CBCS), The Cancer Genome Atlas (TCGA), The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), and The Sweden Cancerome Analysis Network - Breast (SCAN-B).

## Materials and Methods

### Study populations and datasets

The Cancer Genome Atlas (TCGA) is a large collaboration aimed at conducting standardized molecular profiling of over 30 cancer types and has been described in detail elsewhere (22). We downloaded legacy clinical, RNA sequencing (RNA-Seq), and somatic alteration data for primary breast tumors in TCGA using NCI Genomic Data Commons (GDC, <https://portal.gdc.cancer.gov/projects/TCGA-BRCA>) under dbGap accession phs000178.v1.p1. HRD scores, including loss of heterozygosity (LOH; ref. 23), large-scale transitions (LST; ref. 24), and number of sub-chromosomal regions with allelic imbalance extending to the telomere (NtAI; ref. 25), were extracted from a previous analysis by Knijnenburg and colleagues (26).

The Carolina Breast Cancer Study (CBCS) is a population-based study of black and non-black (98% caucasian, referred to as white) women from 44 counties of central and eastern North Carolina conducted in three phases; study details and sampling schema have been described previously (27–30). According to the CBCS study design, cases were women ages 20 to 74 years, diagnosed with a first primary invasive breast cancer, and identified via rapid case ascertainment. Black and younger women (age <50) were oversampled. Race was determined by self-report. Tumor characteristics for cases (e.g., tumor size, grade, hormone receptor status, node status, and stage) were abstracted from medical records and pathology reports. Patients who provided informed consent completed a baseline questionnaire regarding personal characteristics, including socioeconomic status, insurance status, health behaviors, and health history, in addition to the collection of patient tumor tissue, blood, and medical records (31, 32). The study was approved by the Office of Human Research Ethics/Institutional Review Board at the University of North Carolina at Chapel Hill, conducted in accordance with U.S. Common Rule. Written informed consent and HIPAA authorization were obtained from each participant.

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) includes approximately 2,000 fresh-frozen primary breast tumors collected from repositories in the United Kingdom and Canada. Patients categorized as DCIS, LCIS, and those with incomplete clinical or pathologic data were excluded (33). Gene expression was measured using Illumina HT-12 v3 microarray panel, and normalization procedures have been described previously (33).

The Sweden Cancerome Analysis Network - Breast (SCAN-B) is a multicenter prospective study of primary invasive breast cancer across in Sweden that began in 2010. Patients undergoing treatment for primary breast cancer at one of seven clinical centers were offered enrollment into the SCAN-B cohort as part of standard

clinical procedure. In addition to postoperative tumor samples, participants provided pre- and postoperative blood samples. Patient demographic and clinical information were collected from the Swedish national cancer quality registry, and expression was measured via RNA-seq (34).

To evaluate the HRD classifier cross-species, we used a set of 354 mouse mammary tumors (GSE3165). Gene expression was normalized, the classifier was applied, and then results were compared with tumor features including strain and subtype from previous literature (35).

### Identification of DNA repair clusters in expression data

We curated a list of 51 DNA repair genes representing regulators of error-prone and error-free DNA repair. Pathways included nucleotide excision repair (NER), Fanconi anemia (FA), mismatch repair (MMR), base excision repair (BER), homologous recombination (HR), translesion synthesis (TLS), nonhomologous and alternative end joining (NHEJ/AEJ), checkpoint, cancer testis antigens (CTA) including *HORMAD1* and *MAGEA4* [pathologic cancer-specific activators of HR and TLS, respectively (36, 37)], and APOBEC cytosine deaminase family (Supplementary Table S1; ref. 38).

For CBCS gene expression profiling, we used an RNA counting method suitable for FFPE-derived RNAs (39, 40). Existing RNA data on 1,800 CBCS specimens included expression of a TP53 pathway signature (21, 41), and both intrinsic subtype and risk of recurrence scores (42). We performed cohort-level normalization as described previously (43) by Removing Unwanted Variation (RUV) with the RUVg function from the RUVseq package (44). Gene expression analysis in TCGA, METABRIC, and SCAN-B were limited to the 51 genes list in Supplementary Table S1 to facilitate comparison with CBCS; additional sensitivity analyses conducted with larger gene panels yielded similar results.

### Unsupervised and supervised predictors of HRD and clinical features

#### Unsupervised analysis

Normalized gene expression values were  $\log_2$ -transformed and median centered using Cluster 3.0 (45), and consensus clustering (46) was used to discern patterns of tumors based on their DNA repair gene expression. Stability and robustness of clusters was assessed with Silhouette width (47) and SigClust (Supplementary Table S2; refs. 48, 49). Expression data were visualized by heatmaps using centroid linkage hierarchical clustering in Cluster 3.0 (45) and Complex Heatmap (50) in R. All analyses were performed in R version 4.0.2 (51).

We estimated prevalence differences (PDs), which represent the difference between an index group and a reference group in the proportion of individuals exhibiting a given clinical or demographic characteristic, between the unsupervised homologous recombination deficiency cluster (U-HRD), and all non-HRD samples. PDs and 95% confidence intervals (CI) were estimated using generalized linear models with binomial distributions and identify link functions (52).

#### Supervised analysis

To build the S-HRD classifier, DNA-based HRD score was used to train the data. Briefly, HRD scores in TCGA were taken from Knijnenburg and colleagues (26), calculated by taking the sum of three components of HRD/genome scarring scores: HRD-LOH (23), LST (24), number of sub-chromosomal regions with allelic imbalance extending to the telomere (NtAI; ref. 25). Scores were dichotomized at a cut-off point of 42 in accordance with previous

recommendations (26). We first constructed a matrix of  $\log_2$ -transformed, median-centered gene expression values for the 51 genes in our DNA repair panel. Next, we used classification to nearest centroids (ClANC; ref. 53) using 10-fold cross-validation to select subsets of genes that would distinguish individuals with high HRD from low HRD. For each iteration of gene selection, we used 1–25 genes to predict each HRD status group (2–50 genes total), then assessed sensitivity and specificity of predictions when applied to 90% of data (training set) versus the remaining 10% (test set). We repeated this process 10 times such that each sample would be included in the test set once, and then calculated the average sensitivity and specificity of the classifier in the test and training sets at each number of selected genes. Finally, we chose the optimal number of genes to use in the final model by finding the number of genes at which the F1 score ( $2 \frac{PPV * Sens}{PPV + Sens}$ ) plateaued, averaged between the test and training sets. The final model was used to evaluate sensitivity and specificity versus HRD score and applied to subsequent external datasets without data on continuous HRD scores (CBCS, METABRIC, and SCAN-B). Predicted HRD classes were then used to characterize the prevalence of key clinical features in HRD high and low groups to determine whether similar patterns emerged across all datasets (i.e., TCGA, CBCS, METABRIC, and SCAN-B).

Given our interest in HRD according to ER status and race, we also calculated ER-stratified PDs and 95% CI to determine whether black women in each ER stratum experienced an enrichment of S-HRD high tumors. To determine whether differential misclassification of S-HRD by race biased results, we recapitulated the “true” HRD status of S-HRD high and low tumors by randomly sampling these tumors according to race-specific positive and negative predictive values of the classifier; tumors that were not sampled were assigned a misclassification-corrected HRD status opposite of their initial determination. We repeated this process 1,000 times, fitting the ER-stratified PD model each time to generate a misclassification-corrected distribution of HRD enrichment among black participants. In TCGA, where continuous data were available, we compared continuous HRD score and score components by race and ER status using Wilcoxon ranked sign tests.

To determine whether HRD was associated with somatic alterations in DNA repair genes, we compared the proportions of supervised HRD (S-HRD) high and low individuals in TCGA with RAD51, BRCA1, BRCA2, and TP53 copy number changes and somatic and germline mutations. Of the 420 somatic mutations identified, only 11 (2.6%) were predicted to have low impact or serve as moderators by the Ensembl Variant Effect Predictor. We plotted overall tumor mutation burden against HRD scores and S-HRD to identify associations between HRD and aggregate somatic changes. Finally, we compared the ability of HRD scores and the HRD classifier to distinguish samples with somatic alterations by calculating AUCs for each gene alteration type/gene/classification method trio.

We plotted Kaplan–Meier curves to determine whether HRD classification was associated with 5-year recurrence risk in CBCS. Specifically, we followed women in CBCS with stage I–III breast cancer who received chemotherapy treatment from the time of interview to incidence of recurrence, censoring women at death, loss-to-follow up, or five years, whichever was sooner. Recurrence analyses were restricted to CBCS participants treated with chemotherapy to minimize confounding by indication, given that nearly all stage I–III HRD-high tumors in CBCS were treated with chemotherapy, and stratified by ER status. Within each ER stratum, we compared women with S-HRD high tumors to those with

S-HRD low tumors using Cox proportional hazards models and log-rank tests. We further validated HRD-recurrence associations by running equivalent survival models in METABRIC.

#### Data availability

TCGA data can be downloaded through the Genomic Data Commons at <https://portal.gdc.cancer.gov>, while mouse-specific and SCAN-B expression data are available on Gene Expression Omnibus, accession numbers, and GSE96058. CBCS data are not publicly available due to concerns about patient privacy but may be accessed after submission of a letter of intent and approval from the CBCS steering committee. METABRIC expression data are available after application to the International Cancer Genome Consortium Data Access Committee in the European Genome-Phenome Archive, accession numbers EGAD00010000210 and EGAD00010000211. Analysis code is available at <https://codeocean.com/capsule/4968825/tree/v1>.

## Results

### DNA repair gene expression clusters

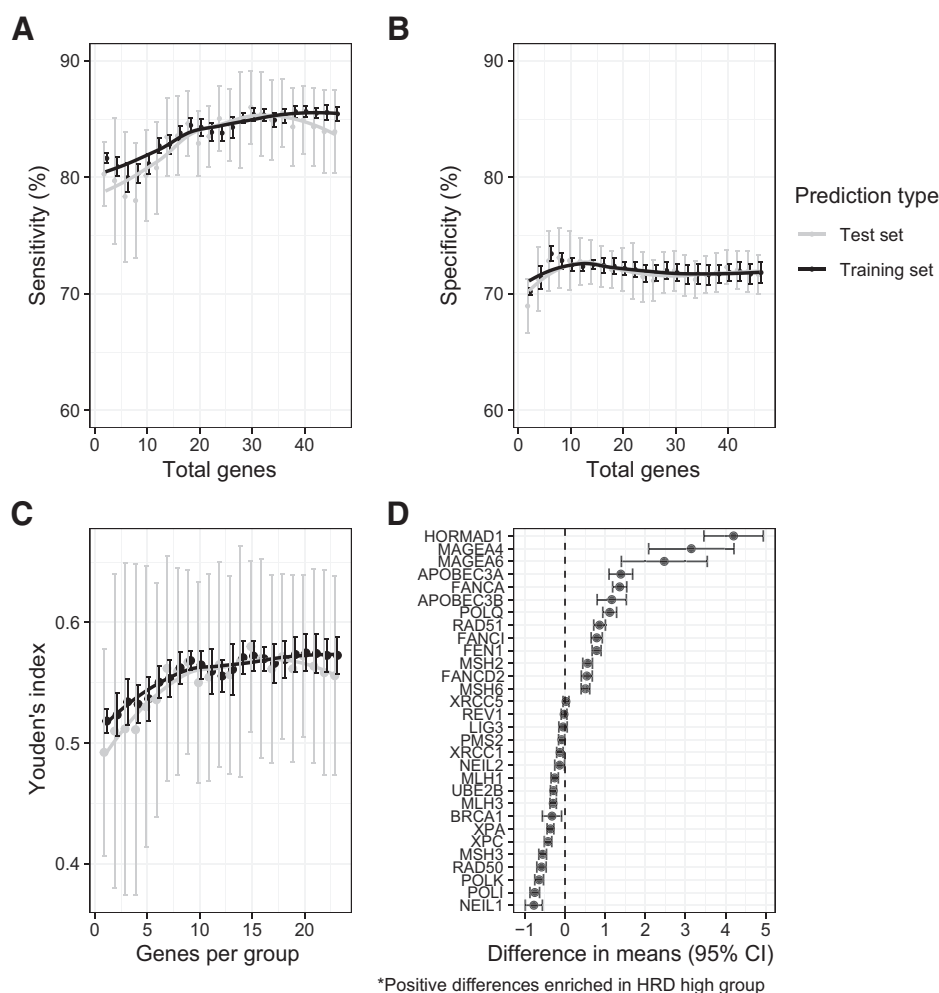
Using unsupervised consensus clustering across 51 genes representing various DNA repair pathways, TCGA breast cancers ( $n = 1,094$ ) clustered into four distinct groups (Fig. 1A). Cluster strength and significance, assessed by Silhouette Width and SigClust, supported four distinct gene expression clusters (Supplementary Table S2). Upon inspection of the clusters, we designated the clusters APOBEC high- or low-expressing according to expression of APOBEC genes among HRD low samples; HRD high or low based on HR/FANC genes, among samples with lower expression of APOBEC3A/3B. The same four groups were detected in CBCS (Fig. 1B), and TCGA unsupervised groups were highly correlated to analogous groups in CBCS by distance to centroid analysis ( $r > 0.7$ ). Supplementary Table S3 shows the distribution of these four groups in TCGA and CBCS. The proportion of tumors with high-HRD was 29.3% in TCGA and 39.7% in CBCS.

HRD is often inferred by a combination, sum score of LOH, telomeric allelic imbalance (NtAI), and LST. To evaluate whether the HR/FA high group was enriched for HRD, we analyzed the distribution of HRD scores by DNA repair group and found that high HRD scores were enriched in the HR/FA high group (Fig. 1C). Therefore, HRD DNA-based status is associated with expression of a specific subset of DNA repair genes, including genes whose products mediate the HR pathway.

In both TCGA and CBCS, the HR/FA high groups had enrichment for basal-like subtype, black race, and TP53 mutant-like RNA status as compared with the HR/FA low groups. When adjusted for TNBC status, the HR/FA high group remained associated with TP53 mutant-like RNA status and black race (Fig. 1D).

We hypothesized that a supervised DNA repair gene expression classifier could be trained to stably recapitulate HRD score. Using ClANC (53), we trained a classifier to distinguish high and low HRD score samples in TCGA. The classifier had approximately 85% sensitivity (Fig. 2A) and 70% specificity (Fig. 2B), with the maximum sensitivity and specificity at 15 genes and 3 genes per group, respectively. Based on Youden index, we trained a model with 15 genes per group (30 total, Fig. 2C, listed in Supplementary Table S1, expression in Fig. 2D). In the full training dataset, the classifier had 84% sensitivity and 73% specificity. The classifier had highest accuracy among





**Figure 2.**

A 30-gene classifier of HRD predicts high HRD status with high sensitivity and specificity. **A**, A gene expression classifier can accurately predict HRD in TCGA breast cancer samples. Mean 10-fold cross validated sensitivity of expression-based classifier using 1–25 genes per group (2–50 genes total) to distinguish samples with high HRD ( $\geq 42$ ) from those with low HRD ( $< 42$ ). ClaNC was used to select genes representative of high and low HRD groups as defined by Knijnenburg and colleagues (26) and predict HRD class. Sensitivity represents the percent of high HRD samples correctly identified as having high HRD. Error bars represent SD across the 10 folds. **B**, Specificity of expression-based classifier using 1–25 genes per group (2–50 genes total) to distinguish samples with high HRD from those with low HRD. Sensitivity represents the percent of low HRD samples correctly identified as having low HRD. Error bars represent SD across the 10 folds. **C**, Youden index (sensitivity + specificity – 1) of expression-based classifier using 1–25 genes per group (2–50 genes total) to distinguish samples with high HRD ( $\geq 42$ ) from those with low HRD ( $< 42$ ). Error bars represent SD across the 10 folds. The optimal number of genes was selected based on maximum Youden index in the test set, which was achieved at 15 genes per group ( $n = 30$  genes total). **D**, Genes in the classifier and their enrichment in HRD high versus low group. Genes are plotted in order of the absolute difference in mean expression between HRD high and low groups, with 95% CIs of the difference also depicted for each gene.

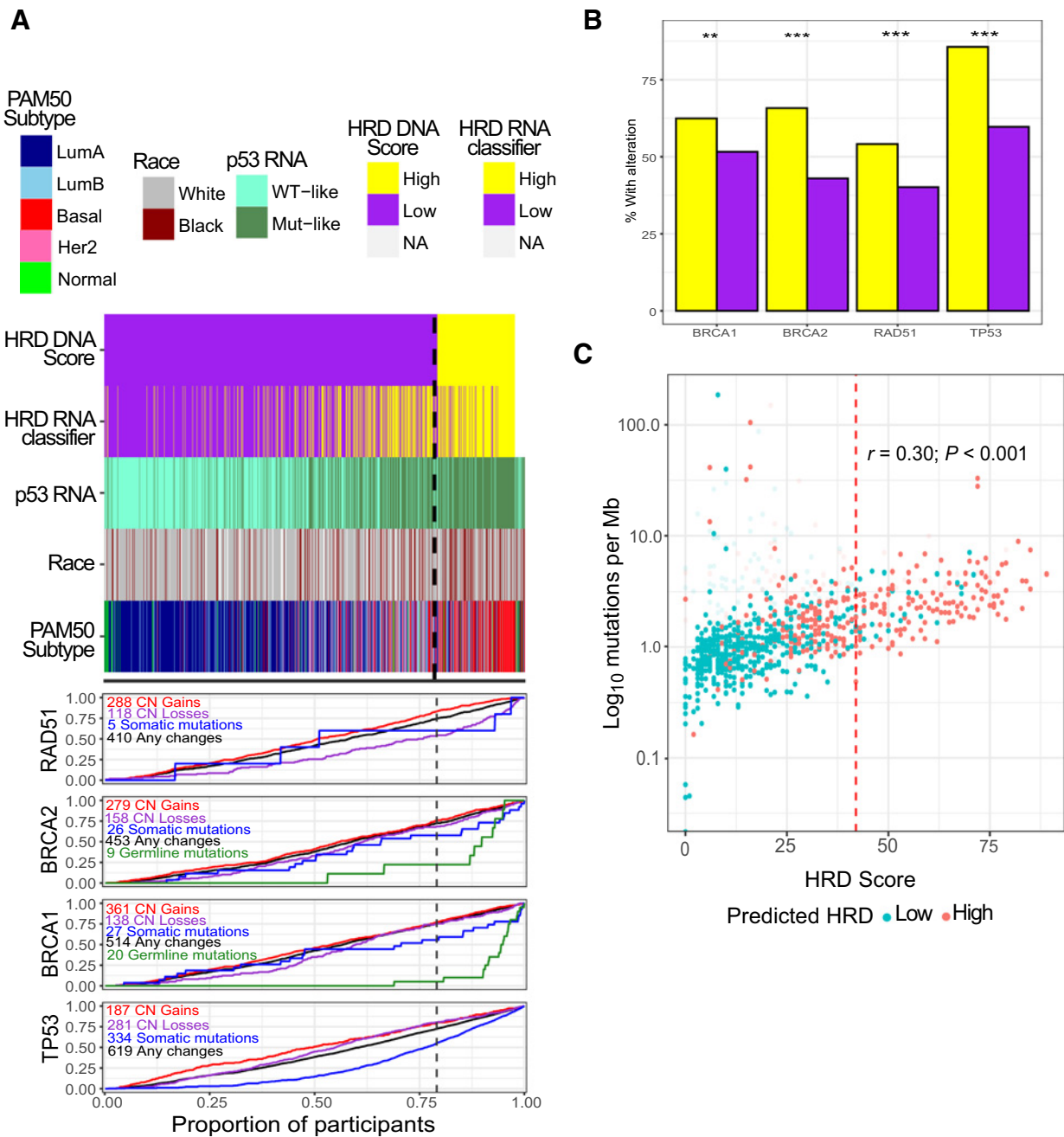
Luminal A subtype (89%), and white race (75%), and highest sensitivity in basal-like (100%) and Her2-enriched (91%) subtypes, and in Black participants (92%; Supplementary Table S4).

Supervised HRD (S-HRD) high samples were enriched for TP53 mutant-like status (PD = 73.3%; 95% CI = 68.9–77.4), black race (PD = 18.2%; 95% CI = 9.6–26.8), and basal-like subtype (PD = 74.5%; 95% CI = 71.0–77.5; Fig. 3A). HRD high samples were also enriched for somatic and germline alterations in RAD51, BRCA1, BRCA2, and TP53 (Fig. 3A and B, with AUC in Supplementary Fig. S1A), with higher rates of SNV and indels in TP53, PALB2, RAD51, BRCA1, BRCA2, and other HRD associated genes (Supplementary Fig. S1B; ref. 54). The overall mutational burden was also higher among S-HRD high samples (median  $\log_{10}$  mutations

per megabase: S-HRD high = 2.05, IQR = 1.33, 3.31; S-HRD low = 1.02, IQR = 0.71, 1.56,  $P < 0.001$ ) and correlated with HRD score (Fig. 3C,  $r = 0.30$ ,  $P < 0.001$ ).

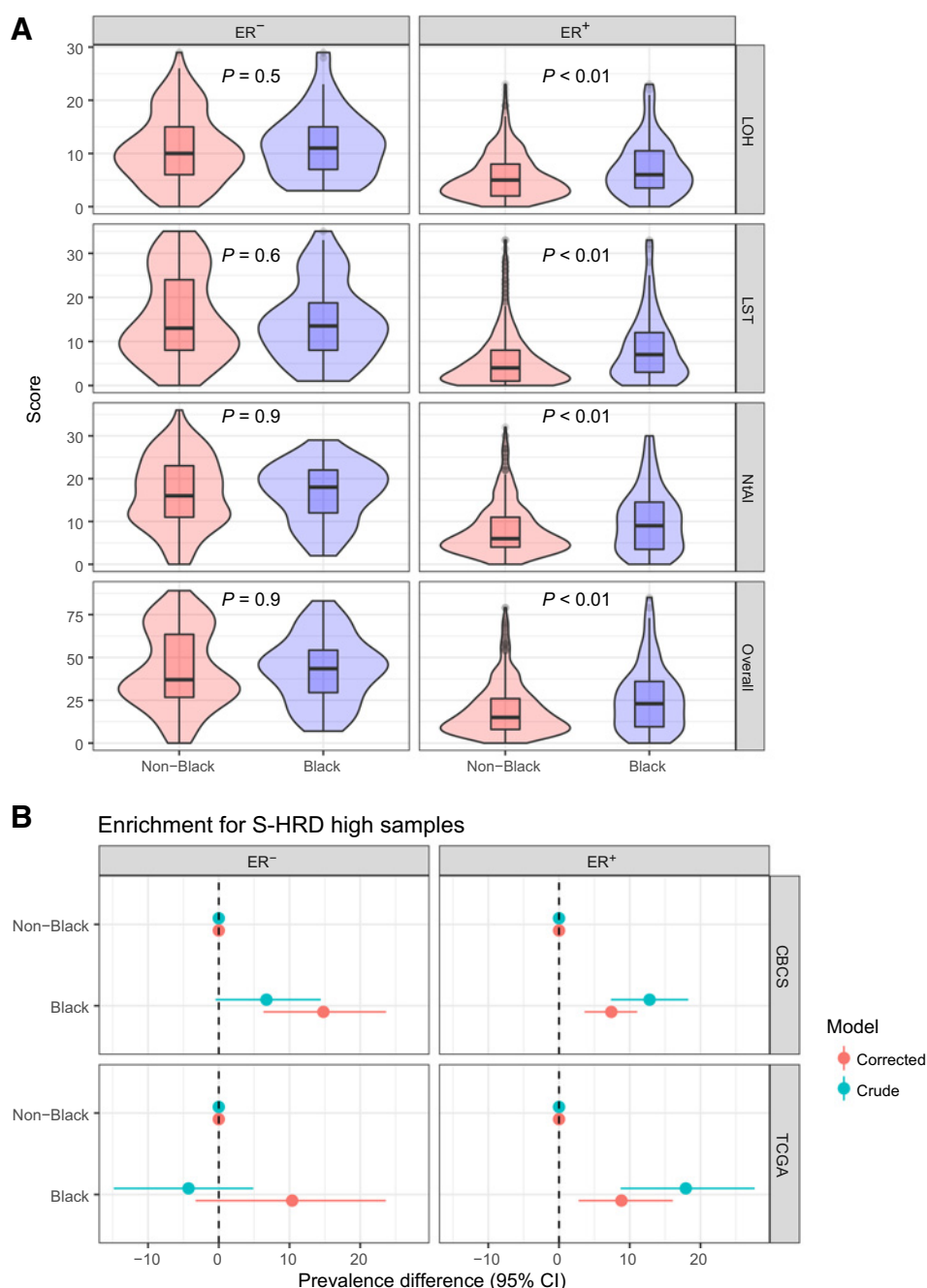
Among ER-negative tumors in TCGA, there were no significant differences in HRD scores or score components (i.e., LOH, LST, NtAI) between black and non-black women. However, among participants with ER<sup>+</sup> disease, black women had significantly higher HRD scores than white women (Fig. 4A). Similar results were observed with S-HRD, although after correcting for misclassification S-HRD was enriched in both ER<sup>+</sup> and ER<sup>–</sup> tumors (Fig. 4B). Applying the S-HRD RNA classifier to the CBCS ( $n = 1,461$ ), we found 813 (55.6%) low and 648 (44.4%) high HRD samples. Similar to TCGA, CBCS HRD high tumors had larger tumor size, and increased frequency of basal-like





**Figure 3.**

The HRD classifier is associated with somatic mutations and mutational burden in TCGA. **A**, Heatmap of clinicopathologic features (above) and somatic and germline alterations (below) in TCGA samples classified as HRD high or low by the HRD RNA classifier. PAM50 subtype, race, TP53 mutant-like status, HRD DNA score, and HRD RNA classifier calls are shown above. Samples missing HRD status are shown in white at far right of all samples. Bottom, cumulative proportions of somatic alterations in selected DNA repair genes. Proportions were calculated as the cumulative number of alterations of a given type divided by the total number of alterations of that type across participants with increasing HRD scores. Red lines indicate copy number (CN) gains, purple lines indicate CN losses, blue lines indicate somatic mutations, green lines indicate germline mutations (only available for BRCA1/2), and black lines indicate any somatic alteration. Vertical dashed line indicates an HRD score of 42, corresponding to the cutoff for high versus low HRD DNA groups. **B**, Percent of HRD high (yellow) and low (purple) samples with somatic alterations in RAD51, BRCA2, BRCA1, and TP53. \*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$  via two sample tests of proportions. **C**, Scatterplot of  $\log_{10}$  mutations per Mb in TCGA samples by HRD DNA score (dashed line is at 42). Dots are colored by HRD RNA calls ( $r = 0.30, P < 0.001$ ). Grayed out dots are samples enriched for the APOBEC mutational signature from COSMIC (19).

**Figure 4.**

HRD is strongly associated with self-reported race in ER<sup>+</sup> cancers, whereas ER<sup>-</sup> cancers show similar HRD phenotypes in black and white women, although misclassification may partially account for this in S-HRD. **A**, Distribution of continuous HRD DNA scores and score components in TCGA ( $n = 875$  individuals with nonmissing HRD scores and self-identification as black or white). **B**, Relative frequency of S-HRD high class, relative to S-HRD low class, according to race and ER status in CBCS ( $n = 485$  ER<sup>-</sup>, 972 ER<sup>+</sup>) and TCGA ( $n = 198$  ER<sup>-</sup>, 630 ER<sup>+</sup>). Misclassification-adjusted bars indicate mean results over 1,000 simulations based on ER and race-specific empirical positive and negative predictive values of the classifier.

subtype and TP53 mutant-like status. HRD high samples were enriched in younger and black participants in CBCS (Supplementary Table S5).

To further explore whether observed patterns between S-HRD and tumor characteristics remained consistent in external datasets, we employed publicly available mouse and human datasets. In METABRIC and SCAN-B, the HRD classifier shared clinical associations observed in TCGA and CBCS, with HRD high samples enriched for basal-like subtype, TP53 mutant-like status, higher grade, larger tumor size, and younger age (Table 1). Using a mouse mammary tumor dataset (35), we measured the proportion of HRD high and low samples using gene expression orthologs. HRD

high class was associated with basal-like and Claudin-low tumor types, and TP53-null or aberrant tumors (Supplementary Fig. S2).

#### HRD classifier and recurrence

Given our interest in the utility of HRD to identify samples at high risk for poor outcomes, we compared the 5-year risk of recurrence in HRD-high versus HRD-low stage I-III chemo-treated CBCS samples, both in ER<sup>-</sup> and ER<sup>+</sup> tumors. As shown in Fig. 5A and B, HRD-high samples had a higher risk of recurrence than the HRD-low samples regardless of ER status, although this difference was statistically significant only in ER<sup>+</sup> tumors. We performed the same analysis in METABRIC and again found that HRD-high samples had a higher risk

**Table 1.** Table of clinical features in low HRD and high HRD classifier samples in METABRIC and SCAN-B datasets.

	METABRIC		SCAN-B	
	HRD Low	HRD High	HRD Low	HRD High
N	1,162	830	1,938	1,471
Age				
≥50 y	953 (82.5)	603 (73.1)	1,637 (84.5)	1,135 (77.2)
<50 y	202 (17.5)	222 (26.9)	301 (15.5)	336 (22.8)
ER Status				
Positive	1,040 (92.3)	459 (56.6)	1,886 (98.8)	1,049 (82.0)
Negative	87 (7.7)	352 (43.4)	23 (1.2)	231 (18.0)
PAM50				
LumA	658 (57.2)	60 (7.3)	1,505 (77.7)	204 (13.9)
LumB	214 (18.6)	274 (33.3)	247 (12.7)	520 (35.4)
Basal	33 (2.9)	296 (35.9)	8 (0.4)	352 (23.9)
Her2	70 (6.1)	170 (20.6)	20 (1.0)	328 (22.3)
Normal	175 (5.0)	24 (2.9)	158 (8.2)	67 (4.6)
Tumor size				
≤20 mm	548 (47.9)	305 (37.5)	1,383 (71.4)	815 (55.4)
>20–≤50 mm	538 (47.1)	464 (57.0)	507 (26.2)	588 (40.0)
>50 mm	57 (5.0)	45 (5.5)	48 (2.5)	68 (4.6)
Grade				
I	158 (14.5)	11 (1.4)	472 (24.5)	33 (2.3)
II	602 (55.1)	170 (21.2)	1,202 (62.5)	391 (27.5)
III	333 (30.5)	622 (77.5)	249 (12.9)	997 (70.2)
Lymph node status				
Negative	659 (57.1)	384 (46.5)	1,231 (65.4)	796 (56.3)
Positive	496 (42.9)	441 (53.5)	652 (34.6)	619 (43.7)

Note: 65 SCAN-B observations missing grade, 111 missing node status. Age, PAM50 subtype, tumor size group, grade, and lymph node status distributions are listed.

of recurrence than the HRD-low samples regardless of ER status, which was significant in both ER<sup>+</sup> and ER<sup>-</sup> tumors (Supplementary Fig. S3).

## Discussion

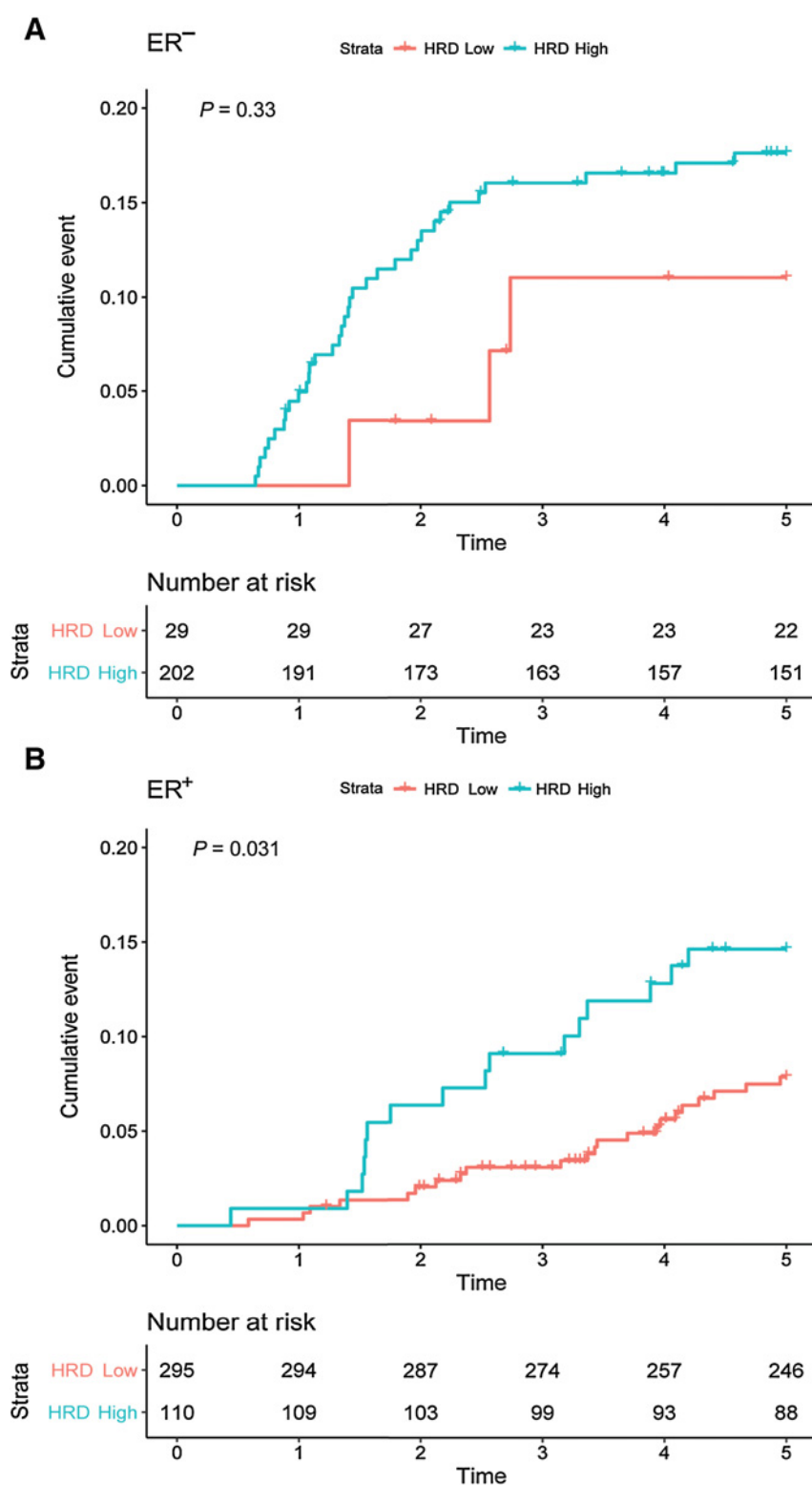
By probing DNA repair gene expression in TCGA and CBCS, we identified RNA-based classifiers of DNA repair imbalances, specifically HRD, that are associated with basal-like subtype, TP53 mutant-like status and that are also present in 25% of ER<sup>+</sup> tumors. Our observation that HRD high tumors were more clinically aggressive and associated with higher hazards for recurrence among ER<sup>+</sup> tumors suggest potential value in prognostication for ER<sup>+</sup> patients who may require treatments beyond endocrine therapy. The potential value of HRD scores has been acknowledged previously (8, 55), but our results indicate that RNA-based classifiers of HRD may be a cost-effective and FFPE-applicable approach for identifying these defects in large population-based studies or clinical specimens. Given that ER<sup>+</sup>, black patients have higher frequency of HRD-high samples and poorer outcomes, understanding the role of HRD in the prognoses of diverse patients is important. However, there is substantial overlap between HRD and other prognostic indicators, including breast cancer subtype and p53 mutation, suggesting that HRD reflects part of a broad molecular vulnerability rather than serving as an independent outcome predictor.

Previous groups have examined the relationship between HRD and clinical outcomes among chemotherapy-treated women. Recently, Boo and colleagues used both a DNA-based HRD signature and a NanoString RNA “BRCA-ness” classifier to predict survival after treatment with adjuvant chemotherapy in patients with TNBC. The authors found modest improvement in survival for patients with HRD

compared with those without, suggesting that tumors with HRD may be more sensitive to chemotherapeutic effects (56). While the limited number of women with HRD who did not receive chemotherapy limited our ability to investigate chemosensitivity directly, our results suggest that ER<sup>+</sup> breast cancers, a group that experiences substantial survival heterogeneity by race, with HRD show a similar profile to ER<sup>-</sup> or triple-negative cancers with HRD (18). Noting that women in historically marginalized populations have often been excluded from precision medicine approaches (57), and that disparities due to systemic racism are often most stark for “treatment-amenable” (e.g., ER<sup>+</sup>) breast cancers (58), the development of a classification method that could be applied to identify ER<sup>+</sup> patients with poorer prognoses could be helpful, although additional analyses will be needed to identify predictors of chemotherapeutic response in these patients.

HRD represents defective homologous recombination, a form of DNA double-strand break (DSB) repair (59). HRD is typically seen in tumors from BRCA1/2 mutation carriers, as the BRCA gene products play an important role in DNA repair via homologous recombination (60). Thus, it may seem counter intuitive that HRD is associated with increased expression of HR genes. However, overexpression of HR/FA genes reflects dysregulation of the pathway, such as higher expression of HR genes (including RAD51) being associated with increased mutation rates (61). The association between HRD and mutational burden has been previously demonstrated in lung (62), breast (63), and other cancers (64). Patterns of DNA repair gene expression can indicate rewiring of DNA repair pathway choice and may reveal specific dependencies and vulnerabilities of cancer cells. HR can be highly error-prone and drives both genome instability and cancer (65–67). In the absence of MMR proteins, the HR pathway may use a mismatched, or “homologous”, donor sequence, which results in



**Figure 5.**

HRD-high classified samples in the CBCS have poorer outcomes among ER<sup>+</sup> tumors. **A**, Cumulative event plot of recurrence over a 5-year period among stage I–III chemo-treated ER<sup>+</sup> participants in CBCS. Blue line indicates HRD high samples and red line indicates HRD low samples. Risk tables are shown below. **B**, Cumulative event plot of recurrence over a 5-year period among stage I–III chemo-treated ER<sup>-</sup> participants in CBCS. Blue line indicates HRD high samples and red line indicates HRD low samples. Risk tables are shown below.

hyperrecombination and causes genome rearrangements (68–71). In addition, pathologic inactivation of some DNA repair mechanisms can be associated with compensatory increases in the expression of other DNA repair mediators. For example, the error-prone DNA polymer-

ase POLQ was found to be elevated at the mRNA level in HR-deficient epithelial ovarian cancers (23). Finally, previous studies have demonstrated that TP53 represses aberrant HR repair (72, 73), and tumors that are enriched for HRD often have mutated TP53 (74).

It is important to consider how DNA repair interacts with immune response. Building evidence suggests that genomic instability may identify tumors that would respond to immunotherapy, and several studies have suggested that different breast cancer subtypes may respond differentially to platinum-based chemotherapies (75, 76). HRD high tumors may be more responsive to platinum-based chemotherapies, especially in high-grade serous ovarian cancer (23, 59, 77, 78). Other groups have examined how HRD impacts pathologic complete response (pCR) after neoadjuvant chemotherapy, with majority of groups showing that HRD status is predictive of pCR (55, 79, 80). There is an unmet need to predict response to chemotherapeutic agents or immunotherapies. Decision tools such as Oncotype DX and Prosigna can help indicate whether chemotherapy is needed, but no widely available genomic assays help to identify specific chemotherapeutic or immunotherapy regimens. Our results suggest that HRD is indeed prognostic, although additional work is needed to confirm its role in treatment response.

There were some limitations of our analysis. Our selected gene list is only a fraction of the possible genes involved in various DNA repair pathways, includes some genes not typical of DNA repair studies (e.g., CTAs), and does not represent all DNA repair variation in breast cancer. Even with a limited gene list, we found consistent patterns of DNA repair gene expression in two divergent breast cancer datasets. We used DNA-based HRD score as the gold standard to train our classifier, but other groups have suggested that HRDetect (81) as an accurate way to measure HRD. We could not evaluate HRDetect in this analysis. Although we used an established cut-off point for determining samples with high HRD scores (8), dichotomization may have introduced further misclassification in our analysis, particularly among samples near the cut-point. More broadly, future studies should consider predictions based on continuous HRD to understand existing variation within HRD categories. We also lacked DNA-based HRD in CBCS, METABRIC, and SCAN-B. Finally, we lacked power to assess the association of HRD with pCR in patients treated with neoadjuvant chemotherapy. Only 59 participants (4.0%) with S-HRD data in CBCS had neoadjuvant chemotherapy and only 29% of the study population had pCR. Given the increased utilization of neoadjuvant chemo, pCR is a priority outcome for future studies of HRD.

In summary, RNA-based assays for DNA repair may address an unmet need in precision medicine, especially for population-based studies with limited genetic and genomic data. It is important to identify how such assays perform in diverse populations and given the sampling structure of CBCS, we were able to document higher rates of HRD in ER-positive breast cancer among black women, suggesting that this pathway merits further consideration in relation to outcome disparities among ER<sup>+</sup> patients.

### Authors' Disclosures

S.C. Van Alsten reports grants from National Cancer Institute during the conduct of the study. G.P. Gupta reports grants from National Cancer Institute during the

conduct of the study, grants from V foundation, and grants from ASTRO/BCRF outside the submitted work. C.M. Perou reports grants from NCI Breast SPORE program P50-CA58223 during the conduct of the study, personal fees from Bioclassifier, LLC. outside the submitted work as well as a patent for U.S. Patent No. 12,995,459 issued, licensed, and with royalties paid from Bioclassifier. No disclosures were reported by the other authors.

### Authors' Contributions

**A. Walens:** Conceptualization, data curation, formal analysis, supervision, validation, investigation, visualization, methodology, writing—original draft, project administration. **S.C. Van Alsten:** Data curation, formal analysis, validation, investigation, visualization, methodology, writing—original draft, project administration. **L.T. Olsson:** Data curation, investigation, writing—review and editing. **M.A. Smith:** Data curation, methodology, writing—review and editing. **A. Lockhart:** Methodology. **X. Gao:** Data curation. **A.M. Hamilton:** Data curation, methodology, writing—review and editing. **E.L. Kirk:** Resources, data curation. **M.I. Love:** Data curation, methodology, writing—review and editing. **G.P. Gupta:** Supervision, methodology, writing—review and editing. **C.M. Perou:** Data curation, supervision, methodology, writing—review and editing. **C. Vaziri:** Conceptualization, supervision, writing—review and editing. **K.A. Hoadley:** Conceptualization, supervision, funding acquisition, writing—review and editing. **M.A. Troester:** Conceptualization, resources, supervision, funding acquisition, writing—review and editing.

### Acknowledgments

We are grateful to CBCS participants for their generous participation, as well as study staff. The authors would like to acknowledge the UNC-CH BioSpecimen Processing Facility for sample processing, storage, and sample disbursements (<http://bsp.web.unc.edu/>). This work and the CBCS was supported by a grant from UNC Lineberger Comprehensive Cancer Center (Chapel Hill, NC), which is funded by the University Cancer Research Fund of North Carolina, the Susan G. Komen Foundation (OGUNC1202 and TREND21686258 to M.A. Troester), and the NCI of the NIH (P01CA151135, to M.A. Troester), including the NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (P50CA058223, to C.M. Perou, M.A. Troester, and K.A. Hoadley). In addition, this work was supported by R01CA253450 (to M.A. Troester and K.A. Hoadley), F31CA257388 (to A.M. Hamilton), Komen Career Catalyst grant (CCR16376756, to K.A. Hoadley), and UNC-CH Cancer Control Education Program (T32CA057726, to A. Walens and S. Van Alsten). This research recruited participants and/or obtained data with the assistance of Rapid Case Ascertainment, a collaboration between the North Carolina Central Cancer Registry and UNC Lineberger Comprehensive Cancer Center. Rapid Case Ascertainment is supported by a grant from the NCI of the NIH (grant no. P30CA016086). The Pathology Services Core is supported in part by NCI of the NIH Center Core Support Grant (P30CA016080) and the UNC-CH University Cancer Research Fund.

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 Epidemiology, Biomarkers & Prevention Online* (<http://cebp.aacrjournals.org/>).

Received May 19, 2022; revised August 3, 2022; accepted September 14, 2022; published first September 21, 2022.

### References

- Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012; 149:979–93.
- Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 2014;15:585–98.
- Jiang T, Shi W, Wali VB, Pongor LS, Li C, Lau R, et al. Predictors of chemosensitivity in triple negative breast cancer: An integrated genomic analysis. *PLOS Med* 2016;13:e1002193.
- Nik-Zainal S, Morganella S. Mutational signatures in breast cancer: The problem at the DNA level. *Clin Cancer Res* 2017;23:2617–29.

5. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* 2019;51:1450–8.
6. Sharma P, Barlow WE, Godwin AK, Pathak H, Isakova K, Williams D, et al. Impact of homologous recombination deficiency biomarkers on outcomes in patients with triple-negative breast cancer treated with adjuvant doxorubicin and cyclophosphamide (SWOG S9313). *Ann Oncol* 2018;29:654–60.
7. Hoppe MM, Sundar R, Tan DSP, Jeyasekharan AD. Biomarkers for homologous recombination deficiency in cancer. *J Natl Cancer Inst* 2018;110:704–13.
8. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous Recombination Deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin Cancer Res* 2016;22:3764–73.
9. Pohl-Rescigno E, Hauke J, Loibl S, Möbus V, Denkert C, Fasching PA, et al. Association of Germline Variant Status with therapy response in high-risk early-stage breast cancer: a secondary analysis of the GeparOcto Randomized Clinical Trial. *JAMA Oncol* 2020;6:744–8.
10. Liao G, Jiang Z, Yang Y, Zhang C, Jiang M, Zhu J, et al. Combined homologous recombination repair deficiency and immune activation analysis for predicting intensified responses of anthracycline, cyclophosphamide and taxane chemotherapy in triple-negative breast cancer. *BMC Med* 2021;19:190.
11. Burstein HJ, Curigliano G, Thürilmann B, Weber WP, Poortmans P, Regan MM, et al. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. *Ann Oncol* 2021;32:1216–35.
12. Moore J, Wang F, Pal T, Reid S, Cai H, Bailey CE, et al. Oncotype DX risk recurrence score and total mortality for early-stage breast cancer by race/ethnicity. *Cancer Epidemiol Biomarkers Prev* 2022;31:821–30.
13. Davis BA, Aminawung JA, Abu-Khalaf MM, Evans SB, Su K, Mehta R, et al. Racial and ethnic disparities in oncotype DX test receipt in a statewide population-based study. *JNCCN* 2017;15:346–54.
14. Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genet* 2015;8:54.
15. Kang J, D'Andrea AD, Kozono D. A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. *J Natl Cancer Inst* 2012;104:670–81.
16. Anurag M, Punturi N, Hoog J, Bainbridge MN, Ellis MJ, Haricharan S. Comprehensive profiling of DNA repair defects in breast cancer identifies a novel class of endocrine therapy resistance drivers. *Clin Cancer Res* 2018;24:4887–99.
17. Killelea BK, Yang VQ, Wang S-Y, Hayse B, Mougalian S, Horowitz NR, et al. Racial differences in the use and outcome of neoadjuvant chemotherapy for breast cancer: results from the national cancer data base. *J Clin Oncol* 2015;33:4267–76.
18. Mazumder A, Jimenez A, Ellsworth RE, Freedland SJ, George S, Bainbridge MN, et al. The DNA damage repair landscape in Black women with breast cancer. *Ther Adv Med Oncol* 2022;14:17588359221075458.
19. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature* 2020;578:94–101.
20. Hurson AN, Abubakar M, Hamilton AM, Conway K, Hoadley KA, Love MI, et al. TP53 pathway function, estrogen receptor status, and breast cancer risk factors in the carolina breast cancer study. *Cancer Epidemiol Biomarkers Prev* 2022;31:124–31.
21. Troester MA, Herschkowitz JI, Oh DS, He X, Hoadley KA, Barbier CS, et al. Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer* 2006;6:276.
22. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013;45:1113–20.
23. 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.
24. 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.
25. Birkbak 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.
26. 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.
27. Millikan RC, Newman B, Tse C-K, Moorman PG, Conway K, Smith LV, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat* 2008;109:123–39.
28. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006;295:2492–502.
29. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat* 1995;35:51–60.
30. Wheeler SB, Spencer J, Pinheiro LC, Murphy CC, Earp JA, Carey L, et al. Endocrine therapy nonadherence and discontinuation in black and white women. *J Natl Cancer Inst* 2019;111:498–508.
31. Allott EH, Cohen SM, Geradts J, Sun X, Khoury T, Bshara W, et al. Performance of three-biomarker immunohistochemistry for intrinsic breast cancer subtyping in the AMBER Consortium. *Cancer Epidemiol Biomarkers Prev* 2016;25:470–8.
32. Allott EH, Geradts J, Cohen SM, Khoury T, Zirpoli GR, Bshara W, et al. Frequency of breast cancer subtypes among African American women in the AMBER consortium. *Breast Cancer Res* 2018;20:12.
33. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486:346–52.
34. Saal LH, Vallon-Christersson J, Häkkinen J, Heggardt C, Grabau D, Winter C, et al. The Sweden Cancerome Analysis Network - Breast (SCAN-B) Initiative: a large-scale multicenter infrastructure towards implementation of breast cancer genomic analyses in the clinical routine. *Genome Med* 2015;7:20.
35. Pfeufferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, et al. Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol* 2013;14:R125.
36. Gao Y, Mutter-Rottmayer E, Greenwalt AM, Goldfarb D, Yan F, Yang Y, et al. A neomorphic cancer cell-specific role of MAGE-A4 in trans-lesion synthesis. *Nat Commun* 2016;7:12105.
37. Gao Y, Kardos J, Yang Y, Tamir TY, Mutter-Rottmayer E, Weissman B, et al. The Cancer/Testes (CT) Antigen HORMAD1 promotes homologous recombination DNA repair and radioresistance in lung adenocarcinoma cells. *Sci Rep* 2018;8:15304.
38. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet* 2014;46:487–91.
39. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 2008;26:317–25.
40. Malkov VA, Serikawa KA, Balantac N, Watters J, Geiss G, Mashadi-Hossein A, et al. Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter™ Assay System. *BMC Res Notes* 2009;2:80.
41. Williams LA, Butler EN, Sun X, Allott EH, Cohen SM, Fuller AM, et al. TP53 protein levels, RNA-based pathway assessment, and race among invasive breast cancer cases. *NPJ Breast Cancer* 2018;4:1–6.
42. Troester MA, Sun X, Allott EH, Geradts J, Cohen SM, Tse C-K, et al. Racial differences in PAM50 subtypes in the Carolina Breast Cancer Study. *J Natl Cancer Inst* 2018;110:176–82.
43. Bhattacharya A, Hamilton AM, Furberg H, Pietzak E, Purdue MP, Troester MA, et al. An approach for normalization and quality control for NanoString RNA expression data. *Brief Bioinform* 2021;22:bbaa163.
44. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 2014;32:896–902.
45. de Hoon MJL, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics* 2004;20:1453–4.
46. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;26:1572–3.
47. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl* 1987;20:53–65.
48. Liu Y, Hayes DN, Nobel A, Marron JS. Statistical significance of clustering for high-dimension, low-sample size data. *J Am Stat Assoc* 2008;103:1281–93.
49. Huang H, Liu Y, Yuan M, Marron JS. Statistical significance of clustering using soft thresholding. *J Comput Graph Stat* 2015;24:975–93.
50. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32:2847–9.

51. R Core Team. R: a language and environment for statistical computing. Version 4.0.2 [software]. 2020 June [cited 2022 Sep 12]. Available from: <https://cran.r-project.org/bin/windows/base/old/4.0.2/>.
52. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* 2005;162:199–200.
53. Dabney AR. Classification of microarrays to nearest centroids. *Bioinformatics* 2005;21:4148–54.
54. Nguyen L, W. M. Martens J, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. *Nat Commun* 2020;11:5584.
55. Telli ML, Hellyer J, Audeh W, Jensen KC, Bose S, Timms KM, et al. Homologous recombination deficiency (HRD) status predicts response to standard neoadjuvant chemotherapy in patients with triple-negative or BRCA1/2 mutation-associated breast cancer. *Breast Cancer Res Treat* 2018;168:625–30.
56. de Boo LW, Józwiak K, Joensuu H, Lindman H, Lanttia S, Opdam M, et al. Adjuvant capecitabine-containing chemotherapy benefit and homologous recombination deficiency in early-stage triple-negative breast cancer patients. *Br J Cancer* 2022;126:1401–9.
57. Ferryman K, Pitcan M. Fairness in precision medicine. *Data & Society* 2018;1–54.
58. Chen L, Li CI. Racial disparities in breast cancer diagnosis and treatment by hormone receptor and HER2 status. *Cancer Epidemiol Biomarkers Prev* 2015;24:1666–72.
59. Li X, Heyer W-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res* 2008;18:99–113.
60. Powell SN, Kachnic LA. Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene* 2003;22:5784–91.
61. Shammam MA, Shmookler Reis RJ, Koley H, Batchu RB, Li C, Munshi NC. Dysfunctional homologous recombination mediates genomic instability and progression in myeloma. *Blood* 2009;113:2290–7.
62. Jiang Y, Dang S, Yang L, Han Y, Zhang Y, Mu T, et al. Association between homologous recombination deficiency and tumor mutational burden in lung cancer. *J Clin Oncol* 2020;38:e21043.
63. Barroso-Sousa R, Jain E, Cohen O, Kim D, Buendia-Buendia J, Winer E, et al. Prevalence and mutational determinants of high tumor mutation burden in breast cancer. *Ann Oncol* 2020;31:387–94.
64. Liu YL, Selenica P, Zhou Q, Iasonos A, Callahan M, Feit NZ, et al. BRCA mutations, homologous DNA repair deficiency, tumor mutational burden, and response to immune checkpoint inhibition in recurrent ovarian cancer. *JCO Precis Oncol* 2020;665–79.
65. Bishop AJR, Schiestl RH. Homologous recombination and its role in carcinogenesis. *J Biomed Biotechnol* 2002;2:75–85.
66. Bishop AJR, Schiestl RH. Homologous recombination as a mechanism for genome rearrangements: environmental and genetic effects. *Hum Mol Gen* 2000;9:2427–334.
67. Guirouilh-Barbat J, Lambert S, Bertrand P, Lopez BS. Is homologous recombination really an error-free process? *Front Genet* 2014;5:175.
68. Spies M, Fishel R. Mismatch repair during homologous and homeologous recombination. *Cold Spring Harb Perspect Biol* 2015;7:a022657.
69. Honda M, Okuno Y, Hengel SR, Martín-López JV, Cook CP, Amunugama R, et al. Mismatch repair protein hMSH2–hMSH6 recognizes mismatches and forms sliding clamps within a D-loop recombination intermediate. *Proc Natl Acad Sci U S A* 2014;111:E316–25.
70. Datta A, Adjiri A, New L, Crouse GF, Jinks Robertson S. Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccaromyces cerevisiae*. *Mol Cell Bio* 1996;16:1085–93.
71. Sugawara N, Goldfarb T, Studamire B, Alani E, Haber JE. Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. *Proc Natl Acad Sci U S A* 2004;101:9315–20.
72. Sengupta S, Harris CC. p53: traffic cop at the crossroads of DNA repair and recombination. *Nat Rev Mol Cell Biol* 2005;6:44–55.
73. Janz C, Wiesmüller L. Wild-type p53 inhibits replication-associated homologous recombination. *Oncogene* 2002;21:5929–33.
74. Holstege H, Joosse SA, van Oostrom CTM, Nederlof PM, de Vries A, Jonkers J. High incidence of protein-truncating TP53 mutations in BRCA1-related breast cancer. *Cancer Res* 2009;69:3625–33.
75. von Minckwitz G, Martin M. Neoadjuvant treatments for triple-negative breast cancer (TNBC). *Ann Oncol* 2012;23:vi35–39.
76. Uhm JE, Park YH, Yi SY, Cho EY, Choi YL, Lee SJ, et al. Treatment outcomes and clinicopathologic characteristics of triple-negative breast cancer patients who received platinum-containing chemotherapy. *Int J Cancer* 2009;124:1457–62.
77. Trenner A, Sartori AA. Harnessing DNA double-strand break repair for cancer treatment. *Front Oncol* 2019;9:1388.
78. da Cunha Colombo Bonadio RR, Fogace RN, Miranda VC, Diz MDPE. Homologous recombination deficiency in ovarian cancer: a review of its epidemiology and management. *Clinics* 2018;73:e450s.
79. Schabath H, Runz S, Joumaa S, Altevogt P. CD24 affects CXCR4 function in pre-B lymphocytes and breast carcinoma cells. *J Cell Sci* 2006;119:314–25.
80. Kim SJ, Sota Y, Naoi Y, Honma K, Kagara N, Miyake T, et al. Determining homologous recombination deficiency scores with whole exome sequencing and their association with responses to neoadjuvant chemotherapy in breast cancer. *Transl Oncol* 2021;14:100986.
81. Davies H, Glodzik D, Morganello 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.