Genome Medicine

RESEARCH

Open Access

Multi-scale characterisation of homologous recombination deficiency in breast cancer

Daniel H. Jacobson[1,2], Shi Pan[1], Jasmin Fisher[2] and Maria Secrier[1*]

Abstract

Background Homologous recombination is a robust, broadly error-free mechanism of double-strand break repair, and deficiencies lead to PARP inhibitor sensitivity. Patients displaying homologous recombination deficiency can be identified using 'mutational signatures'. However, these patterns are difficult to reliably infer from exome sequencing. Additionally, as mutational signatures are a historical record of mutagenic processes, this limits their utility in describing the current status of a tumour.

Methods We apply two methods for characterising homologous recombination deficiency in breast cancer to explore the features and heterogeneity associated with this phenotype. We develop a likelihood-based method which leverages small insertions and deletions for high-confidence classification of homologous recombination deficiency for exome-sequenced breast cancers. We then use multinomial elastic net regression modelling to develop a transcriptional signature of heterogeneous homologous recombination deficiency. This signature is then applied to single-cell RNA-sequenced breast cancer cohorts enabling analysis of homologous recombination deficiency heterogeneity and differential patterns of tumour microenvironment interactivity.

Results We demonstrate that the inclusion of indel events, even at low levels, improves homologous recombination deficiency classification. Whilst BRCA-positive homologous recombination deficient samples display strong similarities to those harbouring BRCA1/2 defects, they appear to deviate in microenvironmental features such as hypoxic signalling. We then present a 228-gene transcriptional signature which simultaneously characterises homologous recombination deficiency and BRCA1/2-defect status, and is associated with PARP inhibitor response. Finally, we show that this signature is applicable to single-cell transcriptomics data and predict that these cells present a distinct milieu of interactions with their microenvironment compared to their homologous recombination proficient counterparts, typified by a decreased cancer cell response to TNFα signalling.

Conclusions We apply multi-scale approaches to characterise homologous recombination deficiency in breast cancer through the development of mutational and transcriptional signatures. We demonstrate how indels can improve homologous recombination deficiency classification in exome-sequenced breast cancers. Additionally, we demonstrate the heterogeneity of homologous recombination deficiency, especially in relation to BRCA1/2-defect status, and show that indications of this feature can be captured at a single-cell level, enabling further investigations into interactions between DNA repair deficient cells and their tumour microenvironment.

Keywords Homologous recombination deficiency, Mutational signatures, PARP inhibition, Breast cancer, Transcriptional classifier, Single cell, Tumour microenvironment

*Correspondence:
Maria Secrier
m.secrier@ucl.ac.uk
Full list of author information is available at the end of the article

Background

Maintaining genomic integrity is an essential process for ensuring the sustained survival of cancer cells and is enabled via a complex network of pathways forming the DNA damage response (DDR) [1, 2]. Homologous recombination (HR) is a robust method of repairing double-strand breaks, and HR deficiency (HRD) causes the cell to become dependent on alternative repair processes such as non-homologous end joining (NHEJ) and theta-mediated end joining (TMEJ) [3]. These reliances provide therapeutic opportunities to target HRD cells with treatments such as poly ADP-ribose polymerase (PARP) and polymerase θ (Pol θ) inhibitors [4–7]. PARP inhibitors have already been accepted for clinical use for patients with breast and ovarian cancer harbouring mutations in *BRCA1* and *BRCA2* [8–10]. However, a significant proportion of cancer patients appear to display evidence of HRD without harbouring these biomarkers [11–13]. Consequently, identifying patients demonstrating HRD, who may therefore benefit from these therapies, has received close attention.

One method for identifying HRD is using patterns of genomic aberrations known as 'mutational signatures' [14]. These signatures act as markers of prior mutagenic events and features, and identifying them within the cancer genome can highlight the driving forces behind the development of a given tumour. Signatures of HRD were uncovered as early as initial studies of mutational signatures [15] and HRD has since been linked with signatures of single-base substitutions (SBS), small insertions and deletions (indels), and structural variants [16, 17]. Specific classifiers have also been created involving the integration of multiple signatures [11] and individual mutational events [12]. However, these classifiers work most optimally when applied to whole-genome-sequenced (WGS) data and do not perform as well when capturing a reduced representation of the genome, e.g. through exome sequencing, which is a common feature of large-scale cancer genomics datasets such as The Cancer Genome Atlas (TCGA).

Alternatives have included measures of chromosomal instability such as copy number signatures, which have been identified in a range of cancers [18–21] as well as the Myriad HRD index score [22, 23], which integrates multiple large-scale genomic events to provide a broader measure of HRD [24–26]. The Signature Multivariate Analysis (SigMA) computational tool applied a likelihood approach to classify both exomes and targeted panel sequenced data as SBS3-enriched [27], which has been shown to successfully predict olaparib response in breast and ovarian cancers [28]. However, the scope of SigMA has been limited to the SBS3 signature alone as an HRD marker, and the inclusion of HRD-associated indel events offers potential for improved classification.

Whilst mutational signatures of HRD have shown substantial potential for predicting PARP inhibitor sensitivity, one drawback of this method is that, because these signatures result from the accumulation of mutations induced by various carcinogens acting from the very early stages of cancer initiation, they are by definition a feature of the history of a tumour, as opposed to its current state. This is of particular importance when it comes to HRD, as mechanisms of HR revival, such as BRCA1 reversion mutations and loss of 53BP1, lead to the emergence of PARP inhibitor resistance in BRCA1-defective patients [29, 30]. A signature based on the expression profile of a sample would, therefore, be more adept at describing the most recent state of a tumour. Transcriptional signatures have been applied to characterise various features associated with HRD, including BRCA1 loss in patients [31], HR gene knockdown in cell lines [32], PARP inhibitor sensitivity [33], and chromosomal instability [34]. Additionally, the presence of signature SBS3 has been used to develop a gene signature for HRD classification in TNBC [35]. This disregards the divergent consequences of BRCA1 and BRCA2 defects due to their different roles in governing HR function: whilst BRCA2 is intrinsically involved in HR via the recruitment of RAD51C to exposed single-stranded DNA, BRCA1 functions upstream of HR and is involved in determining the choice of repair pathway by inhibiting 53BP1, which drives end protection and therefore NHEJ [36, 37]. A transcriptional signature reflecting this heterogeneity may prevent a skewness towards BRCA1-type HRD, whilst also shedding light on the emergence of HRD in BRCA-positive patients.

In order to perform multi-scale characterisation and explore the heterogeneity of HRD, we present a mutational signature-based classifier of HRD for exome-sequenced breast cancers which we then apply to develop a transcriptional signature of HRD and BRCA1/2 deficiency. We demonstrate that leveraging HRD-associated indel events improves HRD classification in downsampled WGS samples and in characterising BRCA-defective samples from the TCGA-BRCA cohort as HRD. Additionally, whilst BRCA + and BRCA-defective HRD samples are broadly similar regarding standard hallmarks of HRD, BRCA + samples show deviations in mutational and phenotypic features such as a comparative decrease in hypoxia. Using matched RNA sequencing data, we then employ multinomial elastic net logistic regression to develop a 228-gene transcriptional signature which can be used to simultaneously predict BRCA1/2 deficiency and HRD status, and is linked with response to PARP inhibitors in both cell lines and patients from the I-SPY2

trial [38]. Finally, we apply the signature to explore HRD at the cellular level from single-cell RNA sequencing data and demonstrate substantial deviations in patterns of crosstalk with the tumour microenvironment (TME) between HRD and HR-proficient tumour cells. Together, these findings demonstrate the value of multi-scale examination of complex phenotypes like HRD and offer opportunities to improve research into the causes and consequences of such deficiencies in human cancers.

Methods

Data sources

Whole-genome sequencing data from the ICGC-BRCA cohort was obtained from DCC data release 28 [39]. Data from two projects was included: BRCA-UK ($n=45$) and BRCA-EU ($n=569$), resulting in a total of 614 samples. These datasets were used to obtain HRD-specific mutational spectra.

Exome sequencing data from primary breast cancer samples from the TCGA-BRCA cohort ($n=968$) was obtained from the GDC Data Portal (https://portal.gdc.cancer.gov/) using the TCGAbiolinks R package [40]. The HRD mutational classifier was applied to this dataset. Somatic mutation data collated using the Mutect2 pipeline was obtained using the GDCquery() function. Annotation of BRCA-defective samples in TCGA was taken from Valieris et al. [41]. Samples were considered BRCA1/2-, RAD51C-, or PALB2-defective if they were assigned either 'Bi-allelic inactivation' or 'Epigenetic silencing'. HRD index scores as calculated by Myriad for TCGA samples were obtained directly from Marquard et al. [22]. CX3 scores were obtained from Drews et al. [20].

Exome sequencing ($n=186$) and transcriptional profiling ($n=168$) from the SMC cohort of Korean breast cancer patients [42], as well as chromosome arm-level copy number alterations for the TCGA-BRCA cohort, were downloaded from cBioPortal [43]. This dataset was used to validate the exome-based HRD classifier and the transcriptional signature (based on 166 matched exomes and transcriptomes).

FPKM-normalised gene expression data for TCGA was obtained from the GDC Data Portal (https://portal.gdc.cancer.gov/). These data were used to develop the transcriptional signature of HRD. The proliferation/cell cycle arrest capacity of the tumours was calculated from RNA-seq profiles using the quiescence (Q) score defined in Wiecek et al. [44] and was defined as 1-Q, with positive scores indicating a higher proliferative capacity.

Gene expression data and PARP inhibitor sensitivity profiles of breast cancer cell lines ($n=26$) were obtained from the Cancer Cell Line Encyclopaedia (CCLE) [45]. Drug sensitivity was measured using the PRISM metric [46]. Transcriptional profiling and clinical data of patients from the patient arm of the I-SPY2 trial ($n=71$) were collected from the Gene Expression Omnibus (GEO) database, with accession number GSE173839 [38]. These datasets were used to assess the clinical relevance of our transcriptional signature of HRD.

Bulk and single-cell RNA-seq data of 515 cells from 13 samples obtained from Chung et al. was collected from GEO database through the accession number GSE75688 [47]. Treatment-naïve single-cell RNA-seq data of 44,024 cells from 14 breast cancer patients obtained from Qian et al. [48] and 84,854 cells from 31 breast cancer patients obtained from Bassez et al. [49] were downloaded directly from the Lambrechts laboratory website https://lambrechtslab.sites.vib.be/en/team. These datasets were used to inspect HRD and its relation to the microenvironment in single cells. Single-cell RNA-seq profiling from Qian et al. and Bassez et al. were processed using the Seurat R package [50], to extract only cancer cells with between 200 and 6000 unique feature counts and mitochondrial content less than 15%, the expression profiles of which were then log-normalised.

Generation of HRD classifier for exome-sequenced breast cancer samples

Mutational signature contributions in WGS breast cancer samples

Mutational signature analysis of 614 WGS breast cancer samples obtained from the International Cancer Genome Consortium (ICGC) project was conducted using the deconstructSigs R package [51]. SBS and indel signatures from the COSMIC v3.3 database were included if they appeared in > 1% breast cancer samples according to the Pan Cancer Analysis of Whole Genomes (PCAWG) project [17]. SBS and indel contributions were calculated separately, and the results were combined for subsequent clustering analysis.

Formation of HRD-specific mutational spectra

The 614 ICGC breast cancer samples were clustered according to their SBS and indel signature contributions, which were calculated separately using the deconstructSigs R package [51]. Since we intended to use WGS samples to calculate estimated signature contributions within the exome regions, we normalised the mutation profiles to account for the frequency of each mutation type occurring in the exome relative to the whole genome. For SBS profiles, this was done by setting the option tri.counts.method = 'genome2exome' in deconstructSigs. For indel profiles, we used the ICGC breast cancer cohort to count the total frequency of each of the 83 indel mutation types, and estimated the frequency of each within the exome by excluding mutations appearing within

intergenic regions. The indel profiles of each sample were then multiplied by the ratio of the frequency of each mutation type within the exome to the whole genome. Additionally, mutation spectra in exomes may also differ from the rest of the genome on account of transcription-coupled damage and repair, which cannot be accounted for in WGS samples even after factoring for varying triplet frequencies.

Model-based clustering was conducted using the mclust R package using finite mixture modelling [52]. The final classification and optimal number of clusters was selected according to the Bayesian Information Criterion value. Whilst 22 clusters were initially identified via this method, two of these clusters consisted of only one sample each, neither of which displayed discernible features; therefore, these clusters were discarded. The 20 remaining clusters were named according to the most prevalent features, with the seven clusters enriched for SBS3 labelled as 'HRD'.

The mutational spectrum for each cluster was determined by collectively calculating the mean distribution of the 96 SBS and 83 indel mutation types. The result is 20 representative mutation distributions consisting of mutation events, each summing to one.

Application of mutational spectra for HRD classification in TCGA

The frequencies of the 179 mutation events across 986 exome-sequenced breast cancer samples obtained from TCGA was determined using the sigminer R package [19]. The probability of a sample being assigned to a specific cluster given the set of aberrations displayed follows Bayes' theorem as follows:

$$P(cluster_i|S) = \frac{P(S|cluster_i) * P(cluster_i)}{P(S)}$$

where S represents the n mutations forming the mutational profile of the sample, $P(cluster_i)$ is the prior probability of assignment to cluster i, with $i \in [1, 20]$, as estimated from clustering of WGS ICGC samples, and $P(S|cluster_i)$ is the likelihood of S occurring in a sample from cluster i, calculated as:

$$P(S|cluster_i) = \prod_{j=1}^{n} P(s_j|cluster_i)$$

where s_j is the jth mutation, $P(s_j|cluster_i)$ is the probability of s_j within the mean mutational spectrum of $cluster_i$, and the normalising constant $P(S)$ is the sum of likelihoods multiplied by prior probabilities for all 20 clusters:

$$P(S) = \sum_{k=1}^{20} P(S|cluster_k) * P(cluster_k)$$

The overall probability of a sample being HRD was calculated as the sum of the probabilities of a sample appearing across the seven HRD clusters, and samples with a probability of greater than 0.79 were deemed HR-deficient. For BRCA-defect-specific HRD classification, samples are assigned to the specific cluster to which they have the greatest probability of assignment.

Evaluation of the HRD classifier

The success of the classifier was determined by calculating its ability to identify patients with HR gene defects. The characterisation of over 900 TCGA samples for either bi-allelic inactivation or epigenetic silencing of BRCA1, BRCA2, RAD51C or PALB2 by Valieris et al. [41] was used as the truth label/gold standard annotation. The F-scores for each HRD classification method were calculated as:

$$F = \frac{TP}{TP + \frac{1}{2} * (FP + FN)}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. Whilst we expected model sensitivity to be close to 100% (as we assumed that the majority of BRCA-defective samples would be HRD) and did not expect these results for model specificity (due to the known presence of BRCA-positive HRD samples), we aimed to ensure that overclassification of HR-proficient samples as HRD was as limited as possible, and so did not apply a weighted F-score. HRD classification using the HRD index score was done using thresholds of 42 and 63 due to their application in the literature [23, 53].

Simulation analysis

We tested our method's HRD classification performance in a low mutation count context using simulated data where we varied the fraction of indels included in a sample. This also allowed us to understand what impact indels generally have for HRD classification. Categories for simulation analysis were generated using hierarchical clustering applied to the SBS signature contributions calculated for the ICGC-BRCA cohort. Samples were downsampled to mutational burdens of 25, 50, and 100, with the additional constraint that the proportion of indels in the simulated data were set from 0 to 0.5, in steps of 0.05. Each combination of sample sizes and indel proportions was iterated 100 times, and differences in the resulting AUC values for identifying SBS3-enriched samples using

the classifier at each iteration were analysed using Wilcoxon rank-sum testing.

To analyse the effect of altering the indel contributions within the likelihood distributions, we repeated the analysis except instead of constraining the indel proportions, we multiplied the indel contribution to each likelihood by increasing factors: 1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5.

Simulation analysis was repeated using all 20 clusters by downsampling each sample to sizes of 25, 50, and 100, with no constraint on indel proportions, calculating the AUC for correct cluster assignment using the classifier, and conducting 100 iterations of this process.

Mutation enrichment analysis

Mutation enrichment analysis was conducted on 738 genes which have been causally implicated in cancer according to COSMIC as of July 3rd 2023 [54]. Only genes which were mutated in more than 5% of samples were included in the analysis, and their enrichment in the HRD/HR-proficient groups was calculated using Fisher's exact test. Mutated genes under positive selection were identified by dN/dS analysis, which was conducted using the dNdScv R package [55] with default parameters. The analysis was run independently for six groups: all HR-proficient, all HRD, BRCA1-defective, BRCA2-defective, RAD51C-defective, and HRD BRCA+.

Chromosome arm-level enrichment analysis

For all chromosome arms, enrichment of CNAs in HRD breast cancer samples compared to HR-proficient samples were calculated using Fisher's exact tests independently testing gains against non-gains (normal or loss), and losses against non-losses (normal or gain).

Pathway enrichment analysis

Differential activity of 14 signalling pathways was analysed using decoupleR [56], which was applied to RNA-seq counts from the TCGA-BRCA cohort. Lowly expressed genes were removed and the remaining data was VSN-normalised. Differential expression analysis was conducted using the DESeq2 R package [57] and the results of this analysis were fed into the decoupleR R package to estimate pathway activity. Cancer type-specific hypoxia scores, using the Buffa signature [58], were obtained from Bhandari et al. [59].

Generation of a transcriptional signature of HRD

Data preprocessing

Samples from the TCGA-BRCA cohort were selected only if they included exome-sequencing data (and therefore had been assigned as HRD/HR-proficient), and a BRCA-defect label obtained from Valieris et al. [41]. To prevent confounding, samples harbouring defects in RAD51C or PALB2 were excluded. This resulted in 857 samples, of which two-thirds ($n=572$) were assigned as training. Training and testing sets were defined using the createDataPartition() function from the caret R package [60] to ensure equal proportions of each HRD/BRCA-defect group within the two sets.

Expression deconvolution

Expression deconvolution was conducted using the BayesPrism R package, which estimates cell type-specific bulk expression profiles from a single-cell RNA-seq reference dataset. In this case, the Qian et al. [48] dataset was used as a reference dataset. Genes from selected groups, including mitochondrial, ribosomal, and chromosome X and Y genes, were excluded. Following this, protein coding genes only were included. To ensure that the resulting signature would be suitably applicable to single-cell RNA-seq data, genes that were expressed in less than 2% of the cancer cells in the Qian et al. dataset were excluded from further analysis, resulting in 9853 genes for downstream analysis.

Development of the multinomial transcriptional signature

The transcriptional signature was generated by multinomial elastic net regularised logistic regression using the glmnet R package [61]. We performed 1000 iterations of tenfold cross validation using the cv.glmnet() function with type.multinomial='grouped'. Initially, four signatures were created, setting alpha=0.25 or 0.5, and applying or excluding weightings to account for group imbalances. We collated the coefficients provided for all features for each iteration in a model generated using λ = lambda.min being the value of λ which gives the lowest mean cross-validated error. The non-weighted elastic net model, with alpha=0.25, was selected due to its presence within the Qian et al. cohort. The final signature was formed of 228 genes which were assigned non-zero coefficients across all 1000 iterations.

Similarly to Severson et al. [31], the signature was calculated using a nearest centroid method. To create the BRCA1/2-deficiency signature, the TCGA training data was split into its four categories, and a template was created for each group by taking the mean expression of each of the 228 genes across the samples in that category. For new samples, four 'scores' were then created by calculating Pearson's correlation coefficient between the expression profile of the new sample and each of the four templates.

The same procedure was also applied to generate an HRD signature, except only two templates were created relating to 'HRD' and 'HR-proficient'. For a new sample, the Pearson correlation coefficients against the two templates were calculated, and then the correlation with the

'HR-proficient' template was subtracted from the correlation with the 'HRD' template to generate an overall HRD score.

This transcriptional signature was compared against four published signatures: a 230-gene HRD signature developed by Peng et al. [32], a 77-gene BRCA1ness signature developed by Severson et al. [31], a 70-gene signature of chromosomal instability (CIN70) [34], and a 7-gene signature of PARPi sensitivity (PARPi7) [33]. Application of these signatures for comparison was also conducted using the centroid method described above. In the event of a gene in the signature not appearing in a dataset, the gene was removed from the signature and did not contribute to the correlation calculation.

Gene set enrichment analysis

Gene set enrichment analysis was conducted using the pathfindR R package [62] and enrichR R package [63]. For the pathfindR analysis, to provide relevant significance values, an ANOVA was conducted for each gene against the four BRCA-defect groups. The KEGG and Gene Ontology Biological Process gene sets were used, and default inputs were used for the remaining arguments.

Importance analysis using a Graph Neural Network Approach

A Graph Attention Network (GAT) was used to map gene co-expression graphs into an embedding space and to analyse its classification output in order to obtain a classification importance score for each gene, indicative of the extent to which this gene's expression can discriminate HRD and HR-proficient samples. The pipeline consisted of following steps:

(1) A weighted correlation network analysis (WGCNA) [64] method was employed to extract a gene co-expression graph involving all 228 genes in the HRD transcriptional signature across the TCGA-BRCA cohort;
(2) A GAT-based feature extractor was applied to the high dimensional embeddings of the gene co-expression graphs which integrates information from neighbouring domains;
(3) A simple prediction module for the classification task was implemented;
(4) A gradient-based method was used to calculate gene importance scores as post-hoc explanations for model behaviour [65].

This pipeline was implemented based on the Pytorch [66] and Pytorch Geometric library [67]. An Adam optimizer was used with batch size 8 and an initial learning rate of 0.002. Linear learning rate decay and early stopping were applied to avoid overfitting. A threshold of 0.7 was applied to identify 'important' genes for HRD and HR-proficiency classification.

Cell–cell interaction analysis

Differential patterns of cell–cell interactivity within the tumour microenvironment of HRD and HR-proficient cells were analysed using CellphoneDB [68], which was applied to the Qian et al. and Bassez et al. cohorts [48]. Tumour cells were labelled as HRD if they displayed a positive HRD score. CellphoneDB was conducted within a Conda virtual Python environment, with default parameters applied.

Statistical analysis

Groups were compared using two-sided Student's *t* test, Wilcoxon rank-sum test or ANOVA, as appropriate. *P*-values were adjusted for multiple testing where appropriate using the Benjamini–Hochberg method. Graphs were generated using the ggplot2 and ggpubr R packages.

Results

Establishing HRD-associated signature phenotypes in whole-genome-sequenced breast cancers

Prior to studying HRD in exome-sequenced samples, we aimed to develop an initial understanding of the recurring profiles of mutational signatures seen in breast cancers. This was achieved by calculating the contributions of breast cancer-associated SBS and indel signatures in 614 whole-genome-sequenced breast cancers obtained from the International Cancer Genome Consortium (ICGC), and then clustering these profiles to established 'signature phenotypes'. According to finite mixture modelling, the ICGC cohort could be grouped into 20 clusters, seven of which were assigned as 'HRD' due to the enrichment of SBS3 and the indel signatures ID6 and ID8 (Additional file 1: Fig, S1a-b; Additional file 2: Table S1).

As expected, *BRCA*-defective samples were strongly enriched within the seven HRD clusters, as were samples labelled as HRD by either HRDetect [11] or CHORD [12], with all BRCA-defective samples appearing in an HRD cluster (Additional file 1: Fig. S1c). 117/120 (97.5%) of samples with HRDetect scores greater than 0.7 appear in an HRD-associated cluster. Considering samples with known BRCA status, 75/135 (55.6%) of samples within the HRD clusters are BRCA-defective, similar to the 74/120 (61.7%) of samples with HRDetect scores greater than 0.7. This is demonstrative of the fact that, even accounting for defects in a range of HR genes such as *RAD51C* and *PALB2* which would increase the precision of these classifiers, the source of HRD for many samples

is unknown, and therefore the specificity for any HRD classifier is not expected to reach 100%.

Interestingly, BRCA1 and BRCA2-defective samples could also be broadly separated. Whilst this has been demonstrated using CHORD [12], we show that it can also be achieved using mutational signatures. BRCA2-defective samples are enriched in clusters characterised by increased contribution of the ID6 signature, referred here as 'BRCA2-type HRD'. The ID6 signature displays deletions at microhomologous regions flanking double-strand breaks, indicative of high TMEJ activity [69]. Alternatively, BRCA1-defective samples appear in clusters featuring increased ID8 signature contributions, which we call 'BRCA1-type HRD', associated with NHEJ [17]. Since BRCA1 is heavily involved in determining how a double-strand break will be processed, BRCA1-defective samples will naturally rely on non-homologous end joining to be their primary method of DSB repair. Each type of HRD cluster shows high sensitivity for classifying their respective BRCA defect, with BRCA1 and BRCA2 defects being correctly classified with sensitivities of 68.9 and 70.0% respectively (Additional file 1: Fig. S1d). Whilst CHORD shows greater specificity (73.3 and 93.3% respectively), there is concordance between the BRCA-type HRD clustering and CHORD classification (78.3% for BRCA1-type HRD and 72.2% for BRCA2-type HRD). Thus the indel signatures may not only be useful as HRD-associated signatures, but also shed light on the method employed by a sample for tolerating that type of HRD.

Evaluating HRD in exome-sequenced breast cancers

A minimum of 50 mutations are generally believed to be required for reliable mutational signature inference [51]. By this criterion, SBS signature analysis is unsuitable in 558/968 whole-exome-sequenced samples from the TCGA-BRCA cohort, and indel signature analysis is unsuitable in 958/968 (Additional file 1: Fig. S2). In particular, these samples display a median indel load of 3, and a mean of 5.39, with 94 samples harbouring zero indel events. There is an opportunity to overcome such limitations when it comes to identifying HRD in exome-sequenced cancers by employing the previously described signature phenotypes in whole cancer genomes: rather than identifying the relevant signatures themselves, the signature profile of a cluster can be predicted instead. To achieve this, we developed a likelihood-based computational method which would enable the assignment to the 20 signature phenotypes without the need to calculate the prevalence of specific signatures (Fig. 1a). For each of the 20 clusters, a mean mutational spectrum was calculated, representing a 'ground-truth'/baseline profile against which new sample spectra can be compared (Additional file 1: Fig. S3). Subsequently, each mutation in the profile of a new sample will influence the likelihood of that sample being assigned to each of the signature phenotypes, with the prior probabilities of assignment to each cluster determined by their size in the ICGC-BRCA cohort.

To initially demonstrate the capacity of this technique and the accuracy gained by considering indel events, we simulated exome-sequenced samples by downsampling whole-genomes from ICGC data to see if our framework would correctly classify the downsampled data according to clusters defined by mutational signatures (see Methods). For these simulations, we based the clustering solely on SBS signatures as a conservative approach that would avoid any bias from indel events which might favour our methodology. We sampled events from each ICGC sample with replacement, constraining them to specific proportions of indel events to demonstrate how their inclusion improved classification. Simulated exomes were set to 25, 50, and 100 mutations. At mutation loads of 50, classification of SBS3 enrichment improved when 5–10% of the simulated mutations were constrained to indel events, indicating that even a small number of indels could improve HRD classification (Fig. 1b). As the simulated mutational load decreased from 100 to 25, a larger proportion of indels was required to enable substantial improvement over using SBS events alone (5% versus 20%, Additional file 1: Fig. S4). The mean and median indel proportions across the TCGA-BRCA cohort are 6.75% and 5.88% respectively, thereby demonstrating that the results of these simulations align with real-world features and that the inclusion of indel events improves HRD classification.

Given the demonstrated and known importance of indel events for HRD classification, we sought to understand whether increasing the weights of indels within the likelihood distributions would improve classification. The above analysis was repeated in WGS samples downsampled to 50 mutations, and the respective indel proportions within the likelihood distributions were multiplied by factors increasing from 1/5 to 5 (see Methods). We demonstrate that whilst HRD classification is broadly unaffected when the indel likelihood weight is decreased, accuracy decreases significantly once it is increased, thereby showing that overestimating the importance of these indel events will hamper the accuracy of HRD classification (Additional file 1: Fig. S5). Thus, from our analysis we conclude that the optimal HRD classification is attained when indel spectra are equally weighted as the SNV ones.

Finally, we asked if the full set of 20 phenotypes seen in WGS data could be captured at lower mutational loads. To this end, we conducted simulations to test the

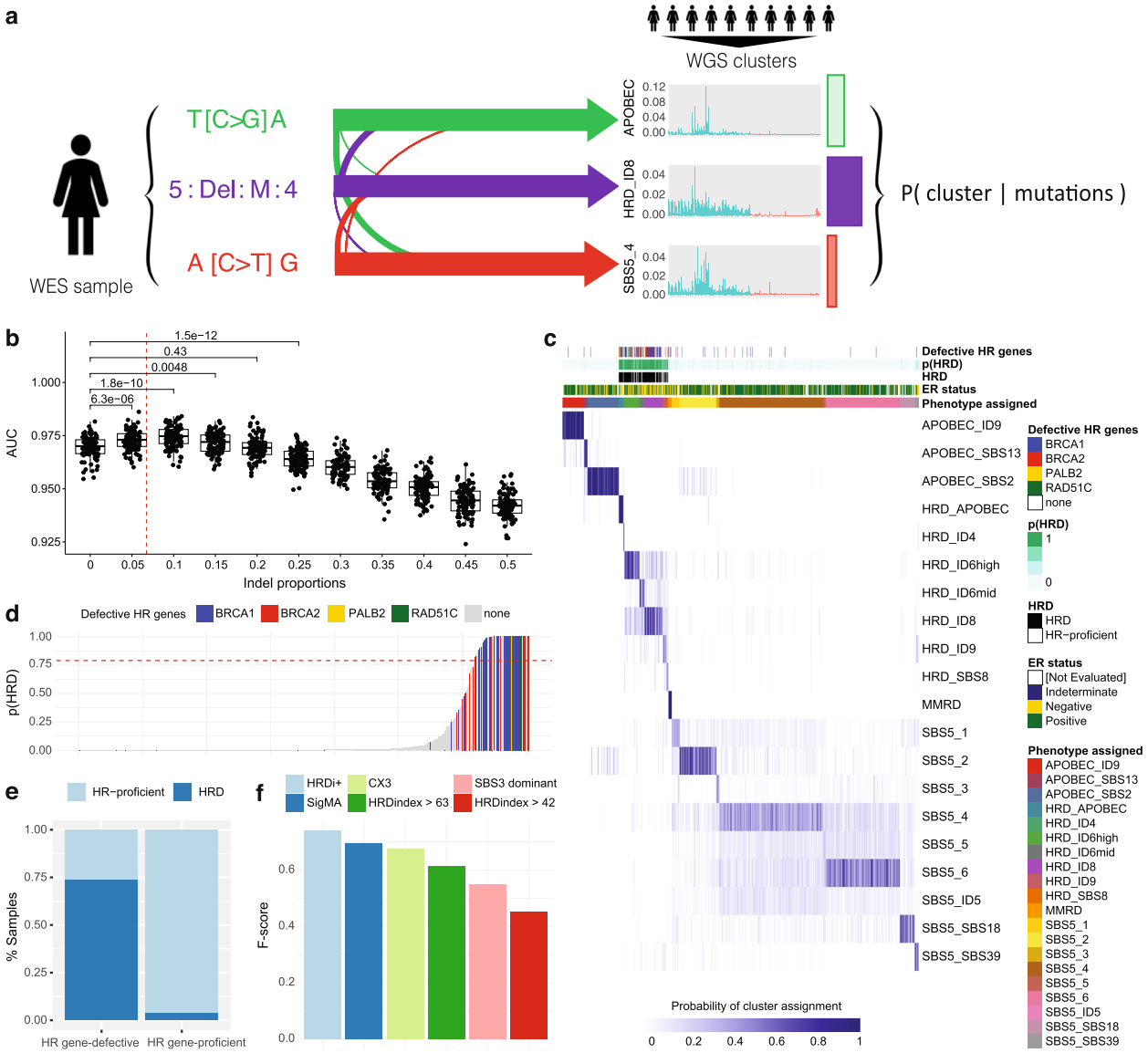

Fig. 1 Evaluating HRD in exome-sequenced breast cancers. **a** Workflow for HRD classification of an exome-sequenced breast cancer sample. Each sample contains a profile of mutations, each of which has a probabilistic association with each of the 20 signature phenotypes, defined by a representative signature profile inferred from WGS data. The mutational profile is collated to calculate the probability of assignment of the respective sample to each of the 20 clusters. **b** Simulation analysis of SBS3-enrichment classification of ICGC samples downsampled to 50 mutational events constrained to varying indel proportions. Adding a small percentage of indels is sufficient to improve classification. AUC = area under the ROC curve for SBS3 enrichment classification. The dotted red line represents the mean proportion of indel events in the TCGA-BRCA cohort. **c** Classification of 968 exome-sequenced breast cancer samples from TCGA. The heat map indicates the probability of each sample (column) being assigned to each signature phenotype (rows). Samples are annotated by ER status and HR gene defects. The value p(HRD) is the sum of probabilities of assignment across the seven HRD-associated phenotypes. The label 'Phenotype assigned' refers to the phenotype to which the respective sample has the highest probability of assignment. **d** Summary of HRD cluster assignment probabilities across the TCGA-BRCA cohort. Samples with a total probability of HRD assignment greater than 0.79 (as shown by the dotted red line) would be assigned as HRD, whereas the rest would be deemed HR-proficient. **e** HRD classification of HR gene-defective and -positive samples in the TCGA-BRCA cohort. 'HRD' refers to samples with a probability of HRD assignment greater than 0.79. **f** F-score comparisons of HRD classifiers for exomes. 'HRDi + ' refers to the classifier developed in this study. The HRD index is presented using cutoffs of 42 and 63

reclassification ability for each phenotype following subsampling (Additional file 1: Fig. S6). This demonstrated that reclassification of specific APOBEC-enrichment clusters was highest at low mutational loads, followed by HRD-enrichment. Additionally, in cases where a sample was misclassified, it was most often assigned to a broader

associated cluster. Thus the APOBEC and HRD phenotypes are reliably identified as such even when the mutation load in a tumour is low or not fully captured by the sequencing technique.

We next applied the likelihood-based classifier to 968 exome-sequenced breast cancer samples obtained from TCGA (Fig. 1c; Additional file 2: Table S2). The samples harbouring defects in HR genes were mainly assigned across the seven HRD-associated clusters. The overall probability of HRD classification was calculated as the sum of probabilities of assignment to the seven HRD-associated clusters, and samples with a classification probability greater than 0.79 were labelled as HRD. This value was selected to maximise the accuracy of the classifier for identifying patients with defects or alterations in HR genes (BRCA1, BRCA2, RAD51C, and PALB2, hereafter termed 'HR gene-defective'), as determined using an F-score (Methods; Additional file 1: Fig. S7a). This method was selected to ensure sufficient sensitivity for classifying samples harbouring known HR gene defects, whilst maximising the confidence in HRD classification of HR gene-positive samples. Overall, HR gene-defective samples from TCGA were labelled as HRD with an AUC of 0.91 (Fig. 1d, Additional file 1: Fig. S7b), 73.4% sensitivity and 74.8% specificity (Fig. 1e).

In general, the accuracy of the classifier is greater when considering ER-negative samples compared to ER-positive ones, likely due to the increased enrichment of HR gene-defective samples within ER-negative samples (31.6% versus 6.79%) (Additional file 1: Fig. S7c-d, Additional file 2: Table S2). It is known that HRD features can vary substantially between cancer types [70], and therefore in some cases redefining specific thresholds based on a given context can improve classification. However, in both cases, the probability threshold of 0.79 performs well for maximising the resulting F-scores.

A similar performance was observed when applying the method in an independent exome-sequencing validation dataset of 186 breast cancer patients from the SMC Korean breast cancer cohort [42]. Here, 13/20 (65%) of BRCA-defective patients were classified as HRD, increasing to 17/20 (85%) when applying a probability cut-off of 0.5 (Additional file 1: Fig. S8).

Our method outperformed other signature-based methods such as SigMA (27), the CX3 copy number signature from Drews et al. [20], or SBS3-based identification alone in terms of both specificity and sensitivity for classifying HR gene-defective samples as HRD (Fig. 1f). It also outperformed the HRD index score, which is based on the levels of loss of heterozygosity, large-scale transitions and telomeric allelic imbalances in a sample (Fig. 1f). We note that the HRD large-scale genomic aberration features required for the Myriad HRD test are often imperfectly called from exome-sequencing data and that efforts have been made to optimise the calling of these features [70]. Therefore, we have used two different thresholds of HRD classification using this HRD index score based on literature-reported cutoffs [53, 71]. Whilst our classifier had a recall in TCGA of only 73.4%, generally lower than some alternative methods proposed, it demonstrated a far superior precision of 74.8%, demonstrating an increased stringency and confidence for HRD classification (Additional file 1: Fig. S9).

Assignment to HRD clusters occurs more frequently for BRCA2-defective samples (93%) compared with BRCA1-defective (72%) and RAD51C-defective (68%) samples (Additional file 1: Fig. S10). However, whilst HRD classification was strong amongst HR gene-defective samples, unlike in the ICGC data, the classifier was unable to assign BRCA-defective samples to BRCA1- or BRCA2-type HRD clusters specifically, with respective sensitivities of 55.0 and 37.0% (Additional file 1: Fig. S10). This is likely due to the scarcity of indel events appearing in exome-sequenced samples.

Hallmarks of HRD

Primary TCGA-BRCA samples labelled as HRD according to the classifier displayed numerous hallmarks associated with DDR deficiencies (Fig. 2), including greater levels of large-scale genomic scarring (Fig. 2a), and high levels of the CX3 copy number signature (Fig. 2b), which has been linked with impaired HRD alongside increased replication stress and impaired damage sensing [20], and which performed particularly well for HR gene-defect HRD classification (Additional file 1: Fig. S11). HRD samples displayed increased expression of POLQ (Fig. 2c), indicating a greater reliance on TMEJ for double-strand break repair [6, 72, 73], as well as increased proliferative capacity (Fig. 2d), which is expected to be associated with HRD [74].

Within TCGA, we see that HRD samples are significantly enriched amongst TNBC samples, with 38.3% of TNBC samples labelled as HRD (with TNBC constituting 40.9% of all HRD samples) compared with 7.1% amongst the remaining samples (Fig. 2e,f), which was also observed in the SMC breast cancer cohort [42]. This aligns with prior information pointing to TNBC samples in TCGA being highly enriched for HR gene defects (43.2% compared with 7.2% amongst the remaining samples). We also classify 55% of the TNBC samples in the SMC cohort as HRD, which is aligned with their findings that 85% of these samples displayed at least some SBS3 signature contribution (Additional file 1: Fig. S8c). Finally, we observed an association between HRD and amplification of MYC (Fig. 2g), which could imply an increase in replication stress [75].

HRD and HR-proficient breast cancers in TCGA were shown to display differential enrichment of mutational drivers of tumorigenesis (Fig. 2h). *TP53* mutations were significantly enriched in HRD samples, in agreement with their common co-occurrence with BRCA1/2 defects [76, 77]. In contrast, the genes *CDH1*, *MAP3K1*, *PIK3CA*, and *GATA3* were more frequently altered in HR-proficient samples. Mutations in *GATA3*, which is involved in normal mammary gland development and has been previously associated with ER positivity, occur frequently in breast cancer, in particular frameshift indels [16, 78, 79] and were strongly enriched in a cohort of Nigerian breast cancer patients [80].

We applied the dN/dS method [55] to identify signals of positive selection for mutations within the HRD and HR-proficient groups (Fig. 2i). Unsurprisingly, mutations in *TP53* and *PIK3CA* were positively selected within both groups, acting as generic drivers in this cancer. Seven genes were positively selected in the HR-proficient group but not HRD. In contrast, *CASP8* and *F5* were positively selected only in HRD samples. Caspase signalling has previously been hypothesised as a driver of cell death following the induction of cGAS–STING and interferon signalling induced by knock-out of BRCA2, indicating that positive selection for *CASP8* mutation could allude to a method of maintaining cell viability in the context of increased chromosomal instablility [81].

BRCA1- and BRCA2-defective samples showed different patterns of positive selection according to dN/dS analysis, with BRCA1-defective and HRD BRCA+ samples showing positive selection only for *TP53*, whilst BRCA2-defective samples only demonstrated this in *PIK3CA* (Fig. 2i) [56].

Due to the chromosomal instability associated with HRD, we also sought to highlight associated copy number aberrations. Chromosome arms were more likely to be enriched for losses than gains in HRD samples, aligning our knowledge of loss of heterozygosity as an HRD signature [26, 82] (Fig. 2j). Only chromosome arm 16p was significantly enriched for both losses in HRD samples, and gains in HR-proficient samples. Notably, the 16p arm carries the *PALB2* gene, which is strongly involved in HR and has been associated with PARPi sensitivity [83–85]. Only three arms were significantly enriched for gains in HRD samples, which were 3q, which carries *POLQ*, 10p, which carries the *DCLRE1C* gene, which encodes for the Artemis protein that is essential for NHEJ activity [86], and 8q, which carries the *SPIDR* and *RAD54B* genes, encoding accessory factors for RAD51C activity during homologous recombination [87, 88], indicating a potential compensatory mechanism in these patients.

Given the variation in mutation selection across HRD samples, we sought to further analyse how BRCA+ HRD samples compare to those harbouring BRCA defects. HRD-BRCA+ samples showed significantly increased levels of HRD hallmarks in comparison to HR-proficient samples (Additional file 1: Fig. S11a). Whilst they showed a slight decrease in CX3 copy number signature contribution compared with BRCA-defective samples, these samples showed no difference in *POLQ* expression or proliferative capacity, indicating that they were displaying a clear HRD phenotype despite their BRCA+ status. Additionally, patients with an HRD classification probability between 0.5 and 0.79 display significantly lower levels of HRD hallmarks compared with those exceeding the threshold, further indicating increased confidence in our HRD classification (Additional file 1: Fig. S11b).

Additionally, we checked the somatic mutation status of HRD HR gene-positive samples in comparison with those carrying HR gene defects. In commonly altered cancer genes, there were no discernible differences between BRCA+ and HR gene-defective samples (Additional file 1: Fig. S12a). Of the 27 HRD HR gene-positive samples, 25 (93%) presented a non-silent mutation in a DDR gene, of which 10 (37%) carried mutations in a double-strand repair gene, according

(See figure on next page.)

Fig. 2 Genomic and transcriptional hallmarks of HRD. Association between HRD status and **a** the Myriad HRD index score, **b** contribution of the CX3 copy number signature, **c** POLQ expression, and **d** a transcriptional measurement of proliferation/cell cycle arrest capacity. **e** Enrichment of HRD across breast cancer subtypes. **f** Enrichment of breast cancer subtypes across HRD status. **g** Association between HRD status and amplification of MYC. **h** Enrichment/depletion of somatic nonsynonymous mutations in key cancer genes in HRD and HR-proficient breast cancer samples. A positive log(odds ratio) indicates enrichment in HRD samples. **i** Positive selection of cancer genes in HR-proficient, all HRD, BRCA1-defective, BRCA2-defective, RAD51C-defective, and HRD BRCA+ breast cancer samples. Circle size indicates the strength of positive selection according to the dN/dS ratio. **j** Comparison of chromosome arm loss and gain events between HRD and HR-proficient breast cancer samples. Positive values indicate enrichment in HRD against HR-proficient samples, whilst the *x* and *y* axes indicate enrichment for chromosome arm gains and losses respectively. **k** Results of differential pathway activity analysis between HRD and HR-proficient breast cancer samples across 14 signalling pathways ordered by the Normalised Enrichment Score (NES). Positive scores indicate pathway enrichment in HRD samples. **l** Comparison of hypoxia scores in the TCGA-BRCA cohort according to the Buffa transcriptional signature across HRD/BRCA-defect categories, split by ER status. *P*-values refer to Wilcoxon testing between each group and the HRD-BRCA+ group, tested across all samples (black) and ER-negative samples only (red)

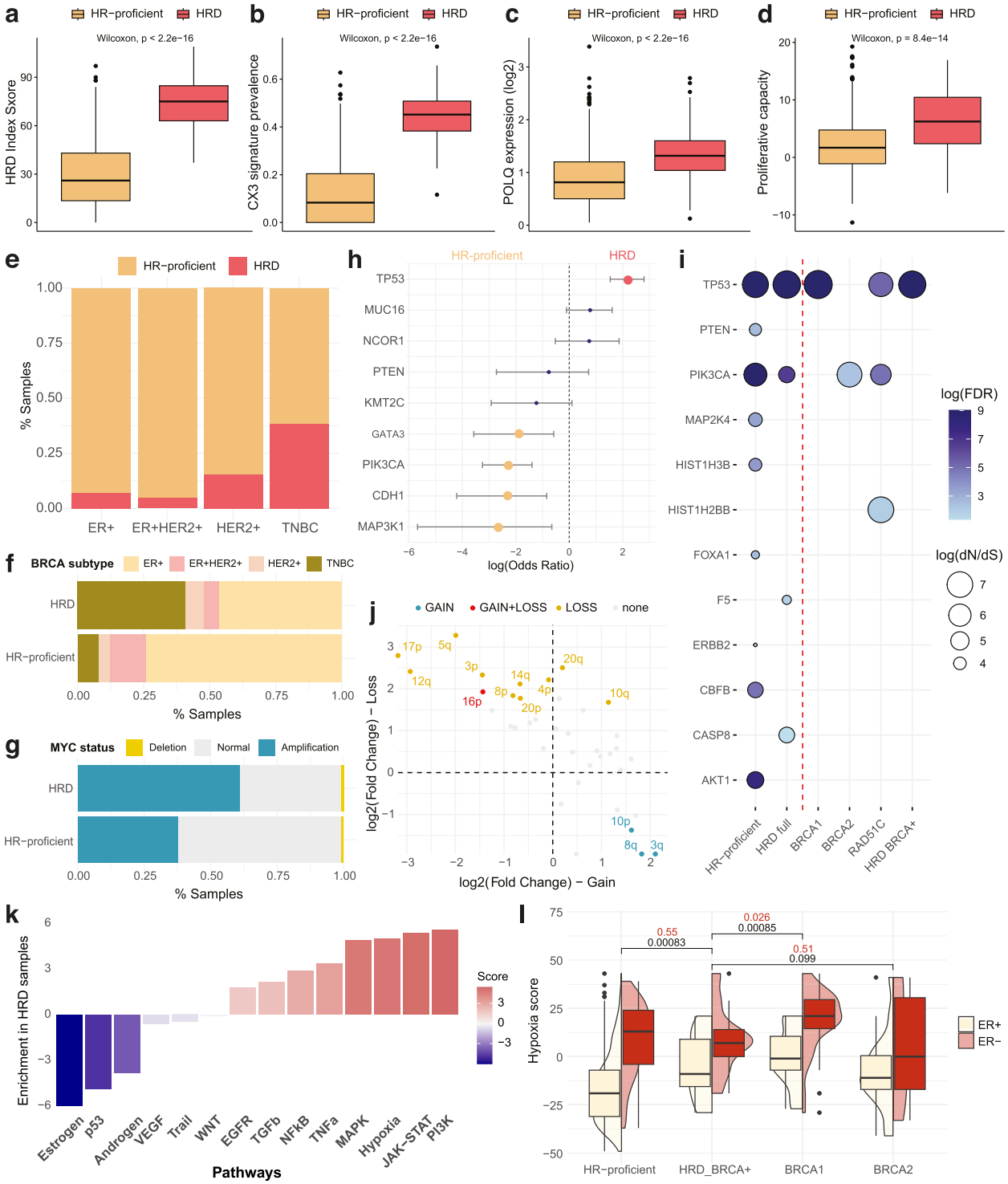

Fig. 2 (See legend on previous page.)

to a curated list of genes associated with DDR [89]. Whilst DDR gene mutations were similarly frequent in BRCA-defective HRD samples (96%), the proportion of double-strand repair-mutated samples, excluding BRCA1 and BRCA2, was significantly higher at 61%. Within HRD BRCA+ samples, 17/27 (63%) carried a *TP53* mutation (similar to 68% amongst HR gene-defective samples), with the next most commonly

mutated DDR genes being *ARID1A* and *MDM4*, a p53 inhibitor, although this was only in 2/27 cases each (Additional file 1: Fig. S12b). Interestingly, mutations in *ARID1A* have previously been associated with PARPi sensitivity [90], indicating a potentially rare cause of HRD. The 16p chromosome arm, carrying the *PALB2* gene, was marginally enriched for gains amongst HRD HR gene-positive samples (Additional file 1: Fig. S12c). However, following multiple testing correction, no genes were enriched for gains or losses within HRD samples depending on HR gene defects (Additional file 1: Fig. S12d).

To gain a broader perspective on the differences driven by BRCA status in HRD samples, we analysed variation in pathway activity using decoupleR [91]. Unsurprisingly, oestrogen and p53 signaling were significantly downregulated in HRD samples (Fig. 2k). We also found that the hypoxia signalling response was substantially upregulated in HRD compared with HR-proficient samples (Fig. 2k). It has previously been demonstrated that BRCA-defective samples from TCGA display increased hypoxia scores compared with BRCA+ samples [92]. Here, we found that HRD HR gene-positive samples also show increased hypoxia scores against HR-proficient samples, although this association disappears when analysing ER-negative samples alone (Fig. 2l). Interestingly, whilst these samples have lower hypoxia scores compared to BRCA1-defective samples, this was not observed in comparison to BRCA2-defective samples. A two-way ANOVA revealed that even after accounting for ER status, both BRCA-defect ($F(2,801) = 11.28$, $p = 1.5e-05$) and HRD status ($F(1,801) = 11.28$, $p = 8.2e-04$) were significantly associated with hypoxia scores (Additional file 2: Table S3). Severe hypoxia has been shown to lead to PARPi sensitivity in HR-proficient tumours [93], and hypoxia has also been shown to inhibit ER expression in breast cancer cells [94], potentially explaining the similar hypoxia levels across BRCA+ ER-negative breast cancers. However, hypoxia as a BRCA-independent mechanism of HRD requires further analysis and experimental validation that is beyond the scope of this study.

Overall, we confirm that the samples classified as HRD via our signature-based method display numerous HRD-associated hallmarks, and demonstrate that HRD and HR-proficient samples show noticeable variation in genomic profiles. Additionally, HRD samples can show variation depending on HR gene-defect status both at mutational as well as signalling activity level, in particular hypoxia, demonstrating physiological deviations which may be reliant on BRCA defects specifically.

Developing a transcriptional signature of BRCA1/2 deficiency

To further explore the functional consequences of HRD, we sought to develop a transcriptional signature reflecting the gene expression profiles of HRD and HR-proficient breast cancers. We aimed to ensure that the signature encompassed the various forms of HRD that could be driven by different factors, such as BRCA1 or BRCA2 defects as well as BRCA-independent mechanisms (HRD-BRCA+). To this end, we trained a multinomial elastic net regression model on the expression profiles from two-thirds of the TCGA-BRCA cohort to distinguish between different forms of HRD or HR proficiency, with the remaining TCGA-BRCA samples used for testing (Fig. 3a, Methods). A multinomial elastic net approach was chosen due to the ability to remove uninformative features, whilst also tolerating correlated variables. Additionally, since we were seeking a cancer cell-specific phenotype, as opposed to a signal of the tumour microenvironment which can confound bulk-sequenced samples, we conducted expression deconvolution using BayesPrism [95] on the training cohort and used the estimated cancer-specific transcriptional profiles for signature development (see Methods). The estimated cell type fractions from BayesPrism significantly correlated with tumour purity estimates [96], indicating reliable cancer cell-specific estimation (Additional file 1: Fig. S13).

The transcriptional signature was generated by conducting 1000 iterations of multinomial elastic net regression with tenfold cross validation and extracting the genes which were included in all 1000 iterations (see Methods). An HRD 'score', as well as four scores representing HR proficiency, BRCA1ness, BRCA2ness, and HRD-BRCA-positivity, were then calculated using a centroid-based approach (see Methods). This procedure was conducted with varying regularisation parameters, as well as with and without observation weights to account for imbalanced groups (see Methods). The final signature, containing 228 genes (Fig. 3b, Additional file 2: Table S4), was selected on account of its optimal capture in single-cell data based on the Qian et al. cohort [48] (Additional file 1: Fig. S14).

The resulting HRD score was adept at classifying samples labelled as HRD or HR-proficient according to the mutational signature-based classifier ($AUC = 0.88$) (Fig. 3c). An advantage of the signature was that whilst BRCA1-defective samples showed the highest HRD scores, BRCA2-defective samples also demonstrated HRD scores greater than HR-proficient samples, demonstrating that this signature can capture overall features of heterogeneous HRD (Fig. 3d). For HRD/HR-proficiency classification, the signature outperformed other

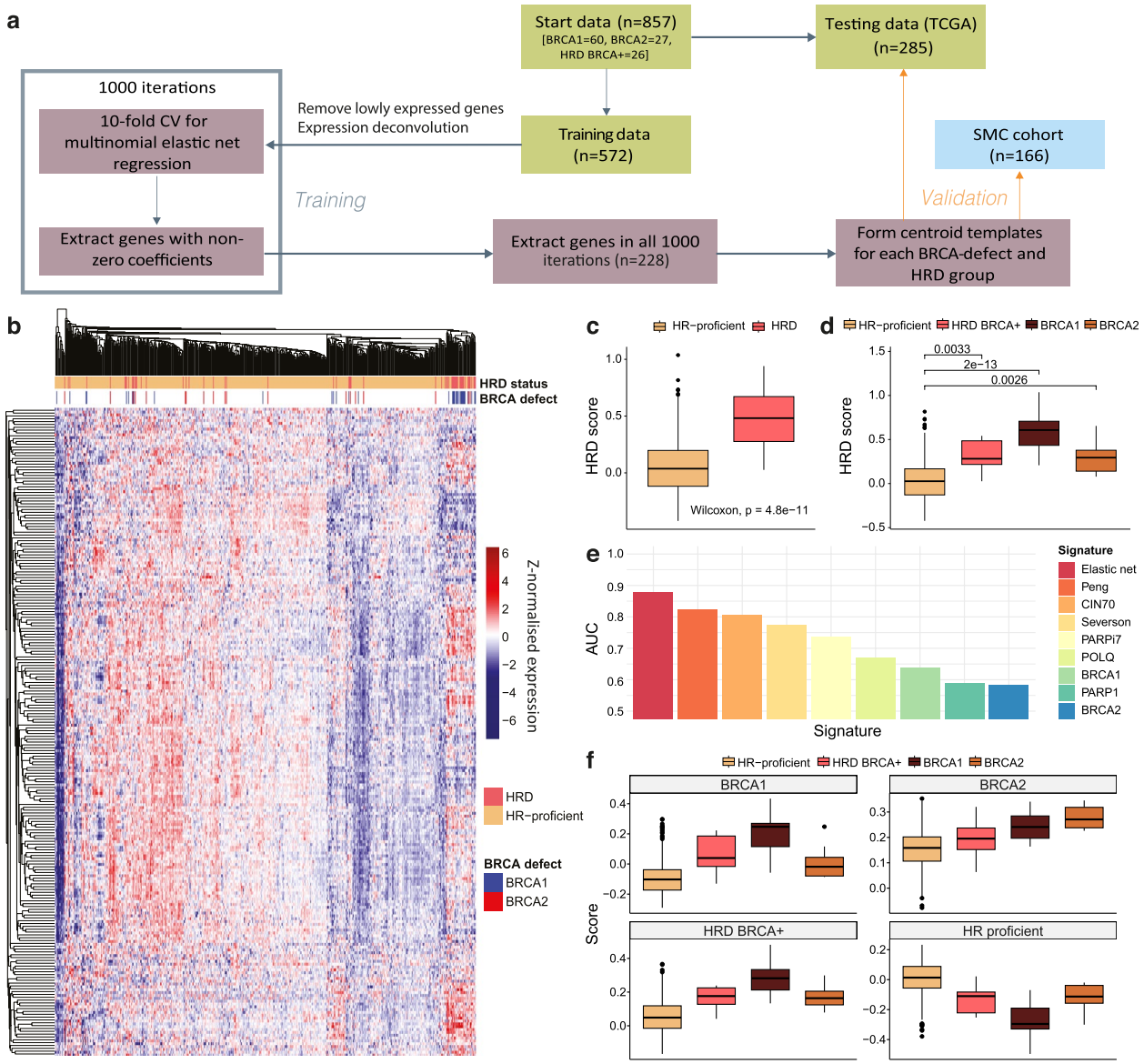

Fig. 3 Development and validation of a BRCA-defect type-specific HRD transcriptional signature. **a** Workflow for transcriptional signature development. Data is split into training and testing cohorts. The training data undergoes expression deconvolution to extract a cancer cell-specific signal, using the Qian et al. single-cell RNA-seq cohort as a reference, and genes that are lowly expressed in this dataset are removed. Processed training data undergoes 1000 iterations of tenfold cross validation of elastic net regression, and a signature is formed from the 228 genes selected in every iteration. Centroid templates are formed for HRD/HR-proficient and BRCA-type HRD groups from the 228 genes across the training cohort, and scores for testing and validation cohorts are calculated by correlating the new sample against each template. **b** Summary of the 228-gene HRD transcriptional signature profiles across the TCGA training set. The HRD status assignment is annotated along with BRCA1/2 defects. **c,d** Comparison of HRD scores calculated using the transcriptional signature between **c** HRD vs HR-proficient and **d** HRD/BRCA-defect groups. **e** Comparison of HRD transcriptional signatures and gene expression markers for predicting HRD status in the TCGA testing set, measured by AUC. 'Elastic net' refers to the 228-gene transcriptional signature presented in this study. 'Peng', 'CIN70', 'Severson', and 'PARPi7' refer to alternative transcriptional signatures as described in the Methods. POLQ, BRCA1, PARP1, and BRCA2 are gene expression markers. **f** Comparison of HRD/BRCA-defect scores across HRD/BRCA-defect groups in the TCGA testing cohort. Each panel corresponds to a specific HRD/BRCA-defect signature, with the *y*-axis representing correlation with the respective centroid model. Each box refers to the samples within the respective group

transcriptional signatures associated with HRD [32], BRCA1ness [31], chromosomal instability (CIN70) [34], and PARP inhibitor sensitivity (PARPi7) [33], as well as gene markers associated with HRD (Fig. 3e).

The signature also showed potential at classifying samples depending on their specific BRCA-defect/HRD status, including when applied to BRCA2-defective samples (Fig. 3f; Additional file 1: Fig. S15). The elastic net signature outperformed all others in characterising BRCA1ness, BRCA2ness and HR proficiency, further demonstrating that, unlike alternative HRD classifiers, using a multinomial approach prevents the resulting signature from skewing away from BRCA2ness, ensuring that the established HRD heterogeneity is captured. Whilst the overall distribution of HRD scores is greater for ER-negative samples in the TCGA testing data regardless of HRD/HR-proficiency status, HRD is also identifiable using the signature after accounting for breast cancer subtype, suggesting that it is not just merely capturing the ER status of the tumours (Additional file 1: Fig. S16).

These results were validated in the SMC breast cancer cohort [42]. The 228-gene signature outperformed alternative methods in HRD classification (AUC = 0.81) and demonstrated adept capacity for BRCA-defect and HRD BRCA+ classification in comparison with alternative methods (Additional file 1: Fig. S17). It is noted that HRD classification capacity is slightly reduced in the SMC cohort which is suspected to be due to the small number of BRCA1-defective patients in the SMC cohort, whilst BRCA1-defective patients dominated the HRD group in TCGA.

According to gene set enrichment analysis, DNA repair processes are dominantly enriched across the signature, as driven by *BRCA1*, *BRCA2*, *TOP3B*, and *FANCI* (Additional file 1: Fig. S18). Intriguingly, the signature was also enriched for genes associated with insulin signalling and glucose transport (*IRS1*, *IRS2*, *CACNA1D*, *SOCS3*, *PRKCZ*), and autophagy and mTOR signalling (*RPTOR*, *RRAGD*, *GABARAP*, *ATP6V1E2*, *ATP6V1C1*).

Key transcriptional contributors to HRD classification

Whilst the 228-gene signature predicts both the HRD and BRCA-defect status, we were interested to explore whether a reduced signature could characterise HRD sufficiently. To achieve this, we employed graph attention networks (GATs) that would help us prioritise genes in the signature which have the greatest contribution to distinguishing HRD and HR-proficient phenotypes (Fig. 4a, Methods). The model makes use of the original genes from the signature and the degree to which they are correlated in their expression within the TCGA-BRCA cohort to classify patients as HRD and or HR-proficient, whilst the correlated gradient information decides which gene sub-modules might drive the classification. Briefly, a patient-specific weighted gene co-expression graph is extracted using weighted correlation network analysis (WGCNA) [64] and these graphs are input into the GAT, which is then trained to distinguish HRD from HR-proficient samples and then selects parts of the graph based on its gradients. The resulting graph neural network showed high accuracy for HRD classification (AUC = 0.90). By ranking the genes using a gradient-based approach for their performance in classifying HRD or HR proficiency, we were then able to highlight the key genes driving correct model prediction.

The analysis highlighted 26 genes which were sufficiently important for classifying the HRD and HR-proficiency groups, of which 15 were predictive of HRD and 11 were predictive of HR proficiency (Fig. 4b,c, Additional file 2: Table S5). A number of these genes have been associated with DDR, including *USP13*, a regulator of replication stress involved in ATR activation via TopBP1 deubiquitination [97, 98], *POLG*, a key regulator of mitochondrial DNA replication and repair [99, 100], and *IP6K2*, a stabiliser of DNA-PKcs and ATM leading to p53 phosphorylation [101, 102]. Additionally, the reduced signature contains two genes encoding zinc finger proteins (*ZNF718* and *ZNF583*) which are involved in both HR and NHEJ [103, 104]. On their own, these 26 genes provide an adept reduced gene signature for capturing HRD, with maintained capacity for HRD classification across BRCA2-defective samples (Additional file 1: Fig. S19).

Testing the signature against sensitivity to PARP inhibitors

To determine the therapeutic relevance of the transcriptional signature of HRD, we next applied the signature to breast cancer cell lines to predict PARP inhibitor sensitivity. The signature was applied to 68 breast cancer cell lines from the Cancer Cell Line Encyclopaedia (CCLE) [45], and matched to PRISM drug sensitivity data from 26 of these cell lines (Additional file 2: Table S6).

Cell lines with increased transcriptional scores for HRD showed increased sensitivity to four different PARP inhibitors, as represented by lower PRISM scores (Fig. 5a). We note that the correlation is marginal, especially regarding olaparib and talazoparib. However, the 228-gene signature still shows a stronger correlation with PARPi sensitivity in these cell lines than any alternative signatures (Additional file 1: Fig. S20). This is likely a result of the signature being developed from bulk-sequenced primary tumour samples, whilst the cell lines will be lacking a microenvironmental component. Whilst we have attempted to account for microenvironmental signals using expression deconvolution, these extracted cancer cell-specific signals will still exist in a broader

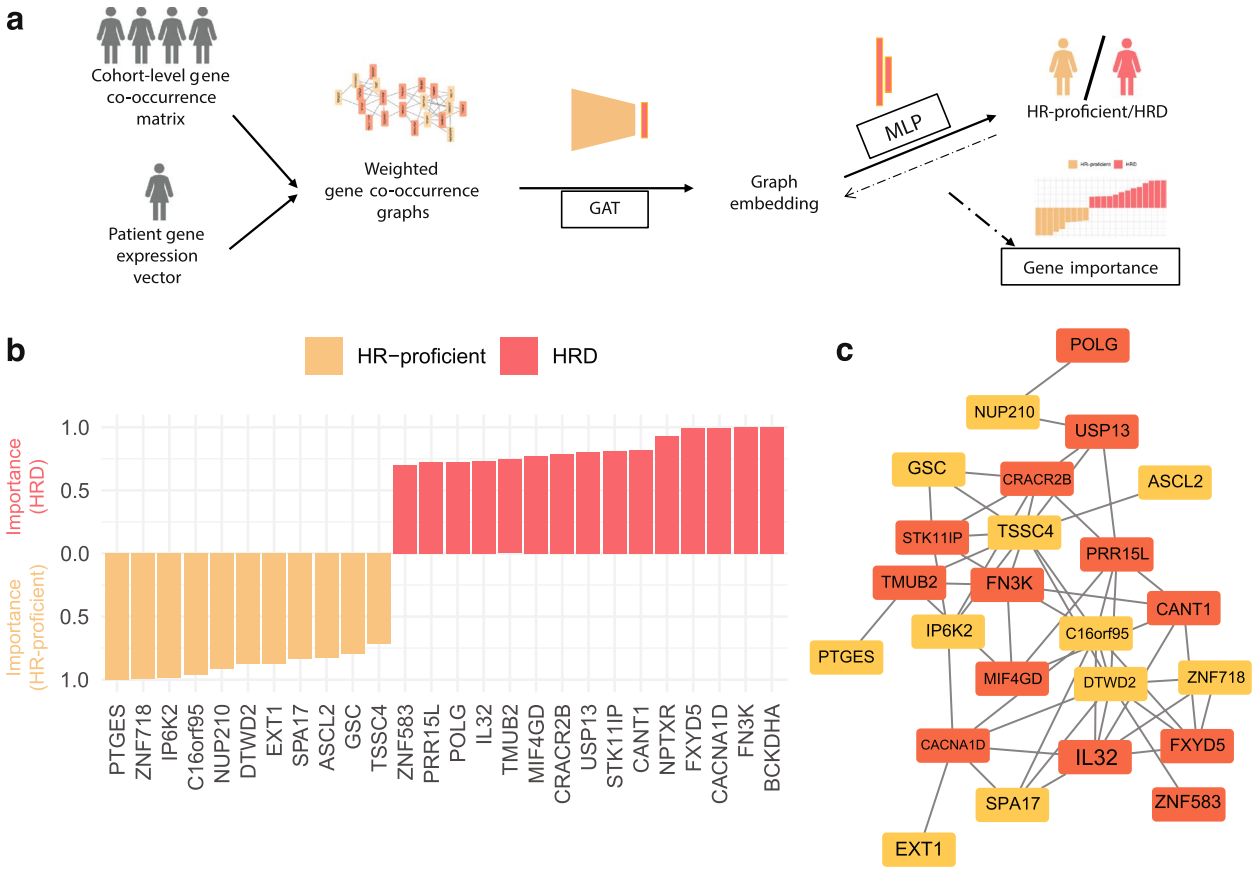

Fig. 4 Graph analysis to determine transcriptional signature drivers. **a** Workflow for graph attention network analysis to classify HRD/HR-proficient TCGA-BRCA patients and determine gene importance. Weighted gene co-expression graphs are built from the gene expression profiles of the TCGA-BRCA cohort, whilst taking into account the patient-level gene expression for the 228 genes in the transcriptional HRD signature. A graph attention network (GAT) is then trained to distinguish HRD and HR-proficient samples using the weighted co-expression graphs as inputs. The output highlights part of the graphs with greater weight in the classification and generates an importance score for each gene. **b** The top-ranked 26 genes in the HRD versus HR-proficiency classification with importance scores greater than 0.7. The colour indicates the phenotype (HRD/HR-proficient) for which the gene is predictive. **c** The co-expression graph of 24 of the 26 highly ranked genes for classification. Genes are connected only if they are co-expressed in the cohort, and genes with no connections have been removed. The colour of the nodes depicts the associated phenotype as in **b**

environmental context which the cell lines will be lacking, hence a resulting transcriptional HRD signal will likely vary significantly.

The HRD score also predicted responses to combined PARP and checkpoint inhibition in breast cancer patients. The signature was applied to 105 HER2-negative, Stage II/III breast cancer patients from the I-SPY2 trial [38]. In this trial, 71 patients were treated with a combination of olaparib and the PD-L1 inhibitor durvalumab, as well as the neoadjuvant chemotherapy taxol, whilst 34 control patients were treated with taxol alone. Amongst the patients in the treatment arm, 29 showed pathologic complete response (pCR), and these patients showed significantly increased HRD scores compared to those who did not display pCR (Fig. 5b). Our HRD score was significantly correlated with their own PARPi7 signature score from the trial but showed a better separation between responders and non-responders (Fig. 5c, Additional file 1: Fig. S21). Overall, this demonstrates that, in capturing a general transcriptional phenotype of HRD, the 228-gene HRD signature is also linked with PARP inhibitor sensitivity, especially in patient samples.

Application of the HRD signature to single cells

When considering HRD in the context of treatment or identification, it is often forgotten that whilst HRD can manifest and cause effects at the level of the whole tumour, it is still an intrinsically cellular phenotype. Since the transcriptional signature was developed with the intention of characterising HRD in terms of

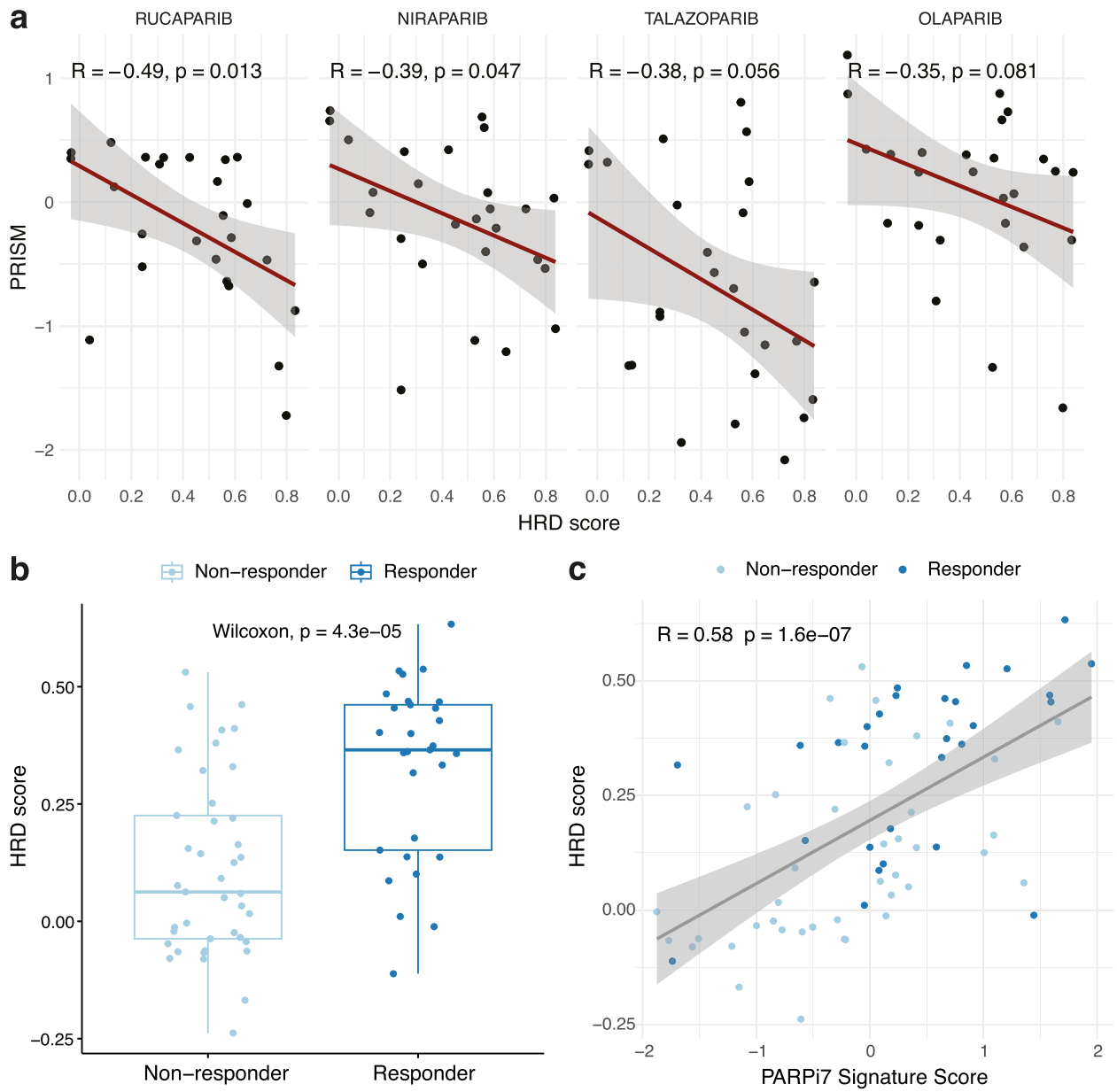

Fig. 5 HRD transcriptional signature is linked with PARP inhibitor sensitivity in breast cancer cell lines and patients. **a** Correlation between the HRD transcriptional scores calculated using transcriptomes from breast cancer cell lines from CCLE and sensitivity to PARP inhibitors evaluated using the PRISM metric. **b** Comparison of HRD transcriptional scores between responders and non-responders to olaparib and durvalumab combination treatment in the I-SPY2 trial. **c** Correlation of HRD transcriptional scores against the PARPi7 signature score calculated by Pusztai et al. (38) in the I-SPY2 treatment arm patients. Responders are more frequently scoring high using our signature compared to PARPi7

tumour-specific signalling, as opposed to that of the microenvironment, this suggested that the signature could be applied to scRNA-seq data, which would provide insight into the distribution of HRD across cells and generate further questions about the varying roles of HRD and HR-proficient tumour cells within the context of the tumour microenvironment.

To investigate whether the signature may be applicable to single-cell data, we first applied it to 11 breast cancer samples with matched bulk and single cell-sequencing data from Chung et al. [47]. We show a good correlation between the bulk HRD scores and the mean HRD scores across the individual tumour cells within each sample (Fig. 6a), providing an indication that the HRD signal

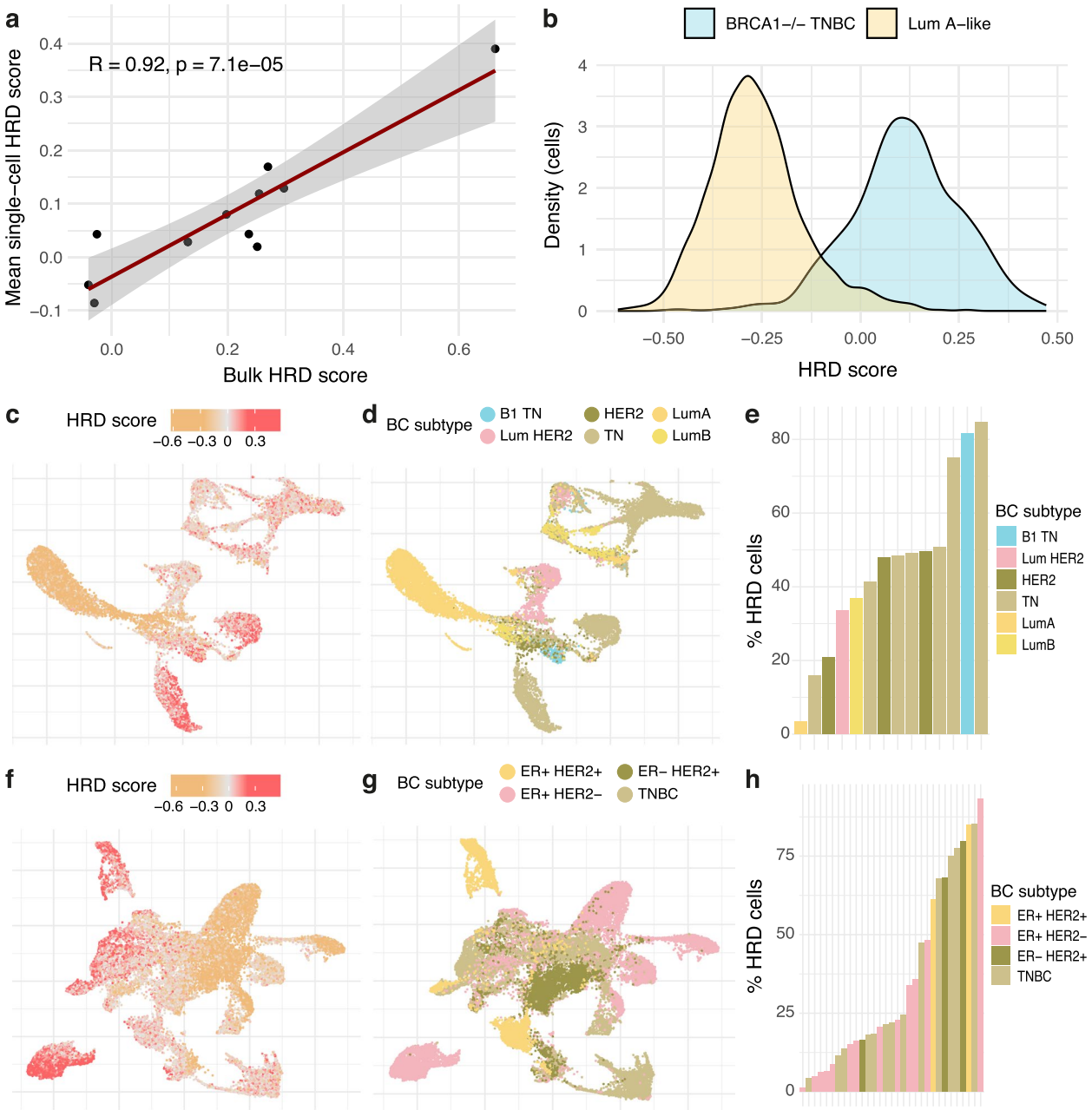

Fig. 6 Transcriptional profiling of HRD in single cell-sequenced breast cancer cells. **a** Correlation of mean HRD transcriptional score across individual cancer cells against matched bulk RNA sequencing from the Chung et al. cohort (47). **b** Distribution of HRD scores across tumour cells from a Stage III BRCA1-defective TNBC sample (sc5rJUQ033) and a Stage II Luminal A sample (sc5rJUQ064) from the Qian et al. cohort (48). **c–e** Profiling of HRD across tumour cells from the Qian et al. cohort as demonstrated by UMAP coordinates labelled by **c** HRD score and **d** breast cancer subtype. **e** The proportion of cells within each sample with HRD scores greater than zero in the Qian et al. cohort. The defined breast cancer subtypes include here are: 'B1 TN' = BRCA1-defective Triple negative, 'Lum HER2' = Luminal-HER2+, 'HER2' = HER2 positive, 'TN' = Triple negative, 'LumA' = Luminal A-like, 'LumB' = Luminal B-like. **f–h** Profiling of HRD across tumour cells from the Bassez et al. cohort (49), similar to **c–e**

captured in bulk sequencing data reflects, on average, the levels seen in single cells.

Following this, we applied the signature to a cohort of 14 single cell-sequenced breast cancers containing over 44,000 cells from Qian et al. [48]. These 14 samples included one BRCA1-defective sample (sc5rJUQ033), which we assumed to be HRD, and one Stage II Luminal A sample (sc5rJUQ064) which we assumed to be

HR-proficient, given the characterisation of this type of breast cancer as slow-proliferating [105, 106]. These two samples display substantially different distributions of HRD scores across the cancer cells in these respective samples, with the Stage II Luminal A sample displaying a distinctively more HR-proficient distribution (Fig. 6b), further demonstrating that a transcriptional signature of HRD can potentially be captured at single-cell resolution.

Across the Qian et al. cohort and 31 treatment-naïve samples obtained from the Bassez et al. cohort [49], on average 10.7% and 11.2% genes from the signature were expressed per cancer cell, respectively (Additional file 1: Fig. S22a-b). The mean proportion of genes in the signature expressed per cell across each sample varied between 6.7% and 17.9% in the Qian et al. cohort (Additional file 1: Fig. S22c) and 6.4% and 23.3% in the Bassez et al. cohort (Additional file 1: Fig. S22d), indicating sufficient capture of the transcriptional signature across the single-cell cohorts.

The cancer cells from these 14 samples displayed intra-sample heterogeneity of HRD scores (Fig. 6c; Additional file 1: Fig. S23a) that matched the clustering by breast cancer subtype (Fig. 6d). Generally, the triple negative and BRCA1-defective samples presented a greater proportion of HRD cells, defined as displaying a transcriptional score greater than zero (Fig. 6e).

Similarly to Qian et al. [48], the tumour cells obtained from Bassez et al. [49] displayed a distinctive gradient of HRD scores (Fig. 6f; Additional file 1: Fig. S23b) in accordance with the clustering by breast cancer subtype (Fig. 6g), and TNBC samples displayed greater proportions of HRD cells in comparison with receptor-positive samples (Fig. 6h). Moreover, we found that, whilst the HRD scores from the TME were consistent across samples and tended to centre closely around zero, these scores varied far more broadly across samples within the cancer cells across both cohorts, indicating that the transcriptional signal of HRD is likely arising strongly from the tumour cells (Additional file 1: Fig. S23c-d).

This further demonstrates the potential provided by this transcriptional signature of HRD to capture this phenotype at single-cell resolution, as well as demonstrating the heterogeneity of HRD levels across individual samples.

Exploration of the HRD tumour microenvironment at a single-cell level

To further explore the activities of HRD and HR-proficient cells within the tumour microenvironment, we applied CellphoneDB [68], an extensive database of ligand-receptor interactions, to observe whether the interactions established between tumour cells and the surrounding immune and stromal cells varied based on HR capacity utilising both the Qian et al. and Bassez et al. cohorts [48]. In both cohorts, cell–cell interactivity profiles were dominated by fibroblast, endothelial, myeloid, and dendritic cells, with cancer cells demonstrating fewer interactions with the TME (Fig. 7a,b). However, across all cell types, HRD cancer cells consistently displayed fewer significant interactions with the TME, in particular as the target of these interactions, than HR-proficient cells (Fig. 7c,d).

Whilst some common interactions were observed, multiple ligand-receptor pairs uniquely mediated the interactions of HR-proficient cells with T-cells across both cohorts (Fig. 7e). In particular, these included interactions involving TNF, TGFb, and prostaglandin E2 signalling across both cohorts and from different cell types (Additional file 1: Fig. S24). These results align with a previous hypothesis suggesting downregulation of TNF signalling as a mechanism of cell survival following BRCA2 deficiency on account of the resulting decrease in caspase-induced apoptosis [81]. Unique TME-HRD cell interaction pairs often involved *LGALS9* signalling, a feature which also re-occurred across cell types and cohorts (Additional file 1: Fig. S24). *LGALS9* has previously been associated with antimetastatic potential in breast cancer [107], potentially indicating a compensation for increased chromosomal instabilility which may drive metastasis [108]. Additionally, expression of the *LGALS9* gene product, Gal-9, is increased following taxane treatment in TNBC, due to nuclear activation of NF-kB, which is also upregulated in HRD breast cancers [109, 110].

This analysis unveils the specific strategies HRD and HR-proficient cells employ in their crosstalk with the TME, highlighting differences in the type and variety of molecular components involved. Generally, the HRD cells appear less 'promiscuous' whilst HR-proficient cells display a wider array of interaction strategies. However, these findings do not inform us on whether HRD cells respond less or more frequently to other cells in their environment, as the individual cell–cell interactions cannot be inferred using this method, and should not be interpreted as such.

Overall, our results highlight a complex pattern of interactions between cancer and non-cancer cells which may be mediated by the DNA damage response and could, in the long term, inform treatment strategies that jointly target HRD tumours and their microenvironments.

Discussion

As the clinical utility and capabilities of personalised treatments increase, it is necessary to ensure that the features predicting positive treatment response can be identified reliably. In this study, we present multi-scale approaches to characterising HRD which can be applied

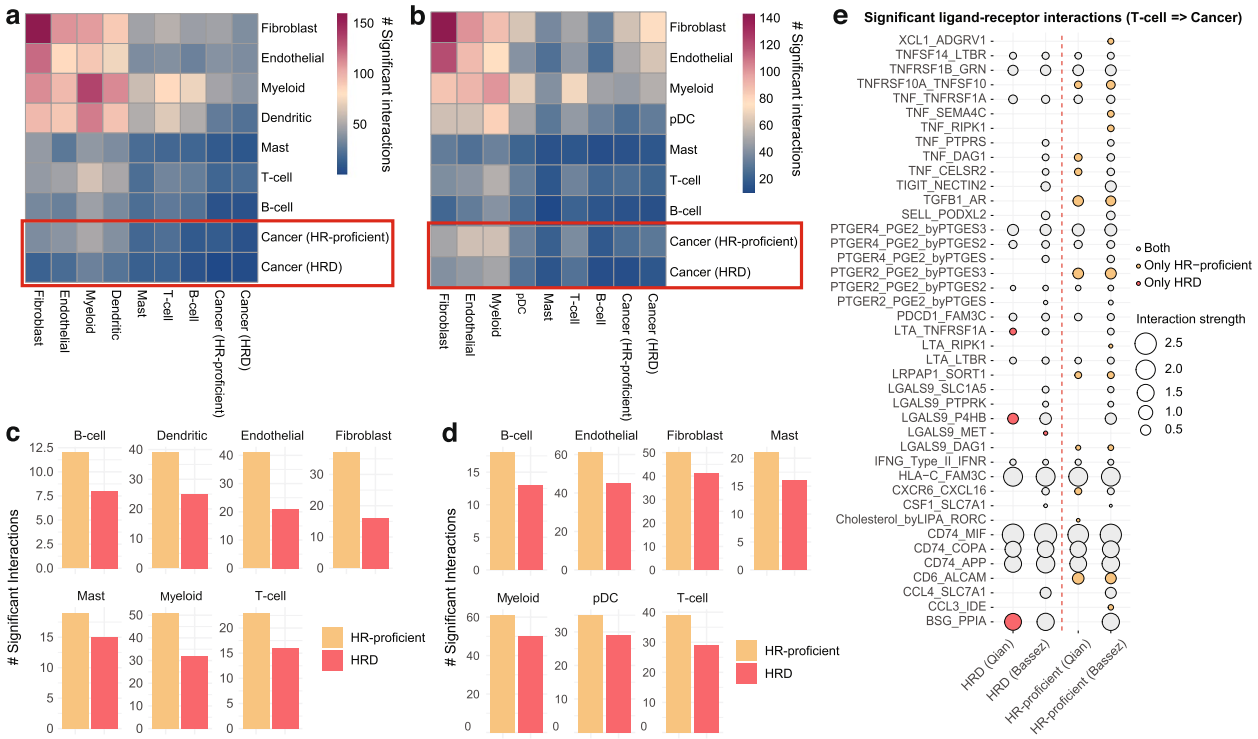

Fig. 7 TME-cancer interactivity across HRD and HR-proficient cancer cells. **a,b** Number of significant ligand-receptor interactions established between cells in the **a** Qian et al. (48) and **b** Bassez et al. (49) cohorts according to CellphoneDB. Cancer cells are labelled as HRD if they have a positive HRD score, HR-proficient otherwise. The *x*-axis refers to cell types as sources, and the *y*-axis refers to cell types as targets. **c,d** The number of significant interactions between TME cell types as sources and cancer cells as targets, separated by HR status, across the **c** Qian et al. and **d** Bassez et al. cohorts. **e** Specific ligand-receptor interactions between T-cells and cancer cells, with cancer cells as the targets, across the Qian et al. and Bassez et al. cohorts. The red circles indicate interactions unique to HRD cells within a given cohort, and the yellow circles indicate interactions unique to HR-proficient cells within a given cohort. Grey circles represent common interactions

in a variety of contexts. We developed a method for high-confidence HRD identification in exome-sequenced breast cancers which incorporates indel events that indicate both the presence of HRD and alternative DSB repair mechanisms employed in the event of HRD. We demonstrate that even small amounts of indels like the ones expected to be seen in WES data improve HRD classification. Furthermore, the HRD group defined by our genomic signatures displays the characteristic features expected of such cancers, including *MYC* amplification and elevated *POLQ* expression. Applying this classifier to the TCGA-BRCA cohort, we then developed a 228-gene transcriptional signature that characterises the heterogeneity of HRD, whilst also correlating with PARP inhibitor sensitivity and displaying the capacity to define HRD at single-cell resolution.

In creating the mutational signature-based classifier, we only considered SBS and indel signatures due to our focus on the HRD phenotype and data availability in TCGA. However, the generalisability of the method is worth highlighting. Copy number signatures could be effectively integrated into this method to improve HRD classification, especially if copy number profiles can be divided into specific features and contexts as demonstrated in previous studies [20, 21]. Furthermore, our method also defines subgroups enriched for alternative mutational processes, including APOBEC cytidine deamination and mismatch repair deficiency. With regard to this, double base substitutions (DBSs) can also be included within the method and may be of use for improved classification of processes with associated DBS signatures, such as mismatch repair deficiencies and tobacco-associated mutagenesis.

We applied a probability threshold of 0.79 for HRD classification using the mutation-based classifier, which was determined through optimising the resulting F-score for identifying patients with HR gene defects. Whilst using this threshold leads to high-confidence classification of HRD in samples which do not harbour HR gene defects (Additional file 1: Fig, S11b), this does lead to a minor but notable decrease in sensitivity. Additionally, we note that these groups may not be distributed identically in the TCGA cohort, as demonstrated by the differences in rates of BRCA defects (13.6% in ICGC compared to

8.99% in TCGA), suggesting that the prior distributions, whilst based on real data, might not be wholly representative. It is non-trivial whether to emphasise sensitivity or specificity when generating an HRD classifier, in which it is known that the specificity should not be 100%. This is because a key difficulty in developing an HRD classifier is the lack of ground truth beyond HR gene defects and, as was used for SigMA and the simulation analysis conducted in this study, SBS3 signature contribution in WGS data. Therefore, we used a balanced F-score which considers both with equal importance. However, we note that different probability thresholds can be applied, as was utilised for SigMA [27].

Mutational signatures have also shown promise for HRD classification in targeted panel sequencing, when presented as a likelihood-based approach as was done through SigMA [27]. Due to the availability of gene panels, this development presents invaluable clinical relevance. Whilst we also employ a likelihood-based approach, owing to the substantially decreased indel loads identified through targeted panel sequencing, it is unlikely that this method would significantly contribute to improved HRD classification, and therefore do not recommend its application to gene panels. Furthermore, it should be noted that our exome classifier's specific clinical utility is limited. Whole-genome sequencing will likely become increasingly available for mutational signature-based diagnostics such as HRDetect, and panel sequencing is already widely applied for identifying gene defects. However, the primary benefit of an exome-based classifier is its application to large-scale genomics resources such as TCGA, enabling further profiling of HRD from a broad range of omics perspectives and new hypotheses generation which these resources enable.

Whilst it is not surprising that our 228-gene transcriptional signature outperforms alternative methods given that it was trained using labels determined by our own mutational HRD classifier, this transcriptional signature also simultaneously captures BRCA1- and BRCA2-specific deficiency phenotypes, highlighting the distinct consequences of loss of function of these two genes. It is worth noting that we additionally capture a group of tumours that display transcriptional profiles closer to those of classically HR-deficient BRCA mutated samples but lacking any BRCA defects. These tumours might be experiencing some level of HRD-like state due to more complex changes across the HR and linked pathways, some of which may be epigenetic or of other nature. Further analyses are needed to shed light into the aetiology of these cancers, and it is likely they are a rather heterogeneous group.

In terms of its relevance in a therapeutic context, we found that our HRD transcriptional signature was more strongly associated with PARP inhibitor sensitivity in patients than in cell lines. Given that the signature was developed using breast cancer patient samples, this was likely to have been the case. Whilst we have attempted to ensure that we are capturing a tumour-cell intrinsic HRD signature by correcting for microenvironmental signals in bulk data, some tumour intrinsic regulation might still be partly environmentally triggered, and this component would not be captured in cell lines which lack this microenvironment. Additionally, due to a broad variety of factors including genetic instability and growth conditions, drug responses across cell lines may be hugely variable [111], which may partially explain the decrease in signature performance when applied to cell lines.

Future developments are likely to focus significantly on PARP inhibitor resistance. Archetypal mechanisms of PARP inhibitor resistance are becoming well established, such as BRCA1 reversion cases, 53BP1 loss following BRCA1 loss, and PARG loss following BRCA2 mutations [112], and features such as reprogramming of cell survival pathways and an increasingly mesenchymal phenotype have been associated with gradual PARP inhibitor resistance [113]. Currently, there are no available datasets demonstrating the effect of HR resurgence on tumour heterogeneity; however, these will provide an invaluable resource for studying HRD moving forward.

Additionally, recent developments in inferring mutagenic processes in single cells have shed light on the driving forces behind the evolution of tumour heterogeneity in TNBC and high-grade serous ovarian carcinoma [114]. In breast cancer, HRD tumours are generally believed to be more immunogenic due to their potential to generate increased mutational loads and neoantigen signalling, which can be exploited for checkpoint inhibition [71, 115, 116] BRCA defects have also been associated with increased immunosurveillance in high-grade serous ovarian cancer, and CellphoneDB was recently applied to highlight a malignant cell population associated with poor prognosis in ovarian cancer, which was associated with immune cell interactions and displayed generally low levels of chromosomal instability [117, 118]. Additionally, CellphoneDB was used to explore human breast cancer immune microenvironments, and specifically highlighted a large number of unique interactions between fibroblasts and endothelial cells, as well as smaller levels of interaction by T-cells and B-cells, as we have also observed, which was attributed to fewer genes being expressed in these cell types [119]. Whilst the TME has been fairly extensively explored in breast cancer bulk datasets and more recently in single cells, our understanding of how the tumour-TME crosstalk is established in the context of HRD or HR proficiency at single cell resolution is much more limited. We show that our

transcriptional signature may be employed to highlight patterns of HRD in single cells, and this paves the way for further explorations into the way that DNA repair deficiencies may influence their microenvironments, and vice versa. Thus, we believe our analysis can pave the way to more detailed interaction studies by highlighting specific strategies of tumour-T-cell coupling in HRD cells which may have relevance to immunotherapy effectiveness. The association between HRD and various facets of the tumour microenvironment is already being mined for therapeutic potential through investigation of joint PARP and immune checkpoint inhibition [38, 120, 121], and chromosomal instability has been shown to elicit an inflammatory response offering further targets for combination treatment [81, 122]. The capacity to investigate differential cell–cell interactivity of HRD and HR-proficient cells may allow for insight into further mechanisms which may be exploited.

Conclusions

We have demonstrated that HRD classification in exome-sequenced breast cancers can be improved by leveraging the presence of HRD-associated indel events, and have shown that mutational and phenotypic profiles of HRD persist regardless of the presence of HR gene defects. These classifications have been used to develop a transcriptional signature which is associated with sensitivity to PARP inhibitors and can be applied to characterise HRD in single-cell RNA sequencing data. In examining the TME crosstalk at single-cell resolution, we demonstrate substantial variation in cell–cell interactivity patterns dictated by the HRD or HR proficiency status of the tumour cells, suggesting distinct pathways mediating immune recognition and/or escape. These findings pave the way to further investigation of the heterogeneity of HRD and HR proficiency both at patient and individual cell level, as well as therapeutic implications.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-023-01239-7.

Additional file 1: Supplementary Material Fig. S1. Establishment of mutational signature phenotypes in whole genome sequenced breast cancers from the ICGC cohort. **Fig. S2.** Single base substitution and indel loads across 968 exome sequenced breast cancers from the TCGA-BRCA cohort. **Fig. S3.** Likelihood distributions of SBS and indel mutation types for each of the 20 signature phenotypes. **Fig. S4.** The impact of varying indel proportions on the reclassification of SBS3-enrichment in whole genome sequenced samples. **Fig. S5.** The impact of varying indel weights within likelihood distributions on reclassifying SBS3-enrichment in whole genome sequenced samples. **Fig. S6.** Analysis of the reclassification ability for the 20 signature phenotypes at lower mutational loads. **Fig. S7.** Determination of the probability threshold for HRD classification. **Fig. S8.** Validation of the mutation classifier in the SMC-BRCA cohort. **Fig. S9.** Precision and recall across HRD classifiers. **Fig. S10.** BRCA type-specific classification in TCGA. **Fig. S11.** Key hallmarks of HRD across varying HRD groups and categories. **Fig. S12.** Cancer driver gene alterations between HRD HR gene-proficient and HR gene-defective samples. **Fig. S13.** Concordance between TME-adjusted cancer cell fractions and tumour purity. **Fig. S14.** Comparisons of single cell expression for multiple signatures generated by regularised logistic regression. **Fig. S15.** Accuracy of transcriptional signatures for predicting BRCA1-defects, BRCA2-defects, BRCA-positive HRD, and HR/BRCA-proficiency within the TCGA-BRCA testing cohort. **Fig. S16.** Comparison of transcriptional HRD scores between HRD and HR-proficient samples within the TCGA-BRCA testing cohort, following separation of samples by ER status. **Fig. S17.** Application of the 228-gene HRD transcriptional signature to the SMC-BRCA validation cohort. **Fig. S18.** Gene Set Enrichment Analysis for the 228-gene HRD transcriptional signature. **Fig. S19.** Performance of the reduced 26-gene HRD transcriptional signature in the TCGA-BRCA testing cohort. **Fig. S20.** Comparison of HRD transcriptional signatures in predicting PARP inhibitor sensitivity in CCLE. **Fig. S21.** Performance of the PARPi7 transcriptional signature for predicting response to olaparib/durvalumab in patients within the treatment arm of the I-SPY2 trial. **Fig. S22.** Signature expression across single-cell RNA-seq breast cancer cohorts. **Fig. S23.** Transcriptional signals of HRD across cancer cells and the tumour microenvironment in single-cell RNA-seq breast cancer cohorts. **Fig. S24.** Significant ligand-receptor interactions between various immune/stromal cells and cancer cells.

Additional file 2: Supplementary Data Table S1. Mutational signature phenotype clustering of the ICGC breast cancer cohort. **Table S2.** HRD classification and signature phenotype assignment of the TCGA-BRCA cohort. **Table S3.** ANOVA analysis of hypoxia levels as a function of ER, BRCA-defect, and HRD status in the TCGA-BRCA cohort. **Table S4.** The 228-gene HRD signature with centroid templates for each group developed using the TCGA-BRCA training cohort. **Table S5.** The reduced 26-gene HRD signature with importance values calculated using a graph attention network. **Table S6.** HRD scores and PRISM values of PARP inhibitor sensitivity for 26 breast cancer cell lines obtained from the Cancer Cell Line Encyclopaedia.

Acknowledgements

Not applicable.

Authors' contributions

MS designed and supervised the study. DHJ developed the mutational and transcriptional classifiers and performed all analyses in bulk and single-cell datasets. SP developed and applied the graph neural network classifier. JF helped co-supervise DHJ and provided valuable feedback on the analyses performed in the study. All authors read and approved the final manuscript.

Funding

DHJ was supported by an MRC DTP grant (MR/N013867/1). MS and SP were supported by a UKRI Future Leaders Fellowship (MR/T042184/1). Work in MS's lab was supported by a BBSRC equipment grant (BB/R01356X/1) and a Wellcome Institutional Strategic Support Fund (204841/Z/16/Z). JF acknowledges the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Availability of data and materials

The datasets analysed during the current study are available at the TCGA Research Network (https://www.cancer.gov/tcga), ICGC (https://dcc.icgc.org/), cBioPortal (https://www.cbioportal.org/), CCLE (https://sites.broadinstitute.org/ccle/), the Lambrechts Lab website (https://lambrechtslab.sites.vib.be/en/data-access) and GEO (https://www.ncbi.nlm.nih.gov/geo/). The following expression datasets from GEO have been employed: GSE173839 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE173839) (38), GSE75688 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75688)(47). All code developed for the purpose of this study can be found at the following repository: https://github.com/secrierlab/MultiscaleHRD [123].

Declarations

Ethics approval and consent to participate

All data employed in this study comply with ethical regulations, with approval and informed consent for collection and sharing already obtained by the relevant consortia where the data were obtained from (TCGA, ICGC).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK. [2]UCL Cancer Institute, University College London, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6BT, UK.

Received: 23 January 2023 Accepted: 26 September 2023
Published online: 02 November 2023

References

1. Ciccia A, Elledge SJ. The DNA Damage Response: Making It Safe to Play with Knives. Molecular Cell. 2010;40:179–204.
2. Lord CJ, Ashworth A. The DNA damage response and cancer therapy. Nature. 2012;481:287–94.
3. Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MIR, et al. Homologous-recombination-deficient tumours are dependent on Polθ -mediated repair. Nature. 2015;518(7538):258–62.
4. O'Connor MJ. Targeting the DNA Damage Response in Cancer. Molecular Cell. 2015;60:547–60.
5. 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(2):123–34.
6. Zatreanu D, Robinson HMR, Alkhatib O, Boursier M, Finch H, Geo L, et al. Polθ inhibitors elicit BRCA-gene synthetic lethality and target PARP inhibitor resistance. Nat Commun. 2021;12(1):3636.
7. Zhou J, Gelot C, Pantelidou C, Li A, Yücel H, Davis RE, et al. A first-in-class polymerase theta inhibitor selectively targets homologous-recombination-deficient tumors. Nat Cancer. 2021;2(6):598–610.
8. Moore K, Colombo N, Scambia G, Kim BG, Oaknin A, Friedlander M, et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. N Engl J Med. 2018;379(26):2495–505.
9. Tutt ANJ, Garber JE, Kaufman B, Viale G, Fumagalli D, Rastogi P, et al. Adjuvant Olaparib for Patients with BRCA1 - or BRCA2 -Mutated Breast Cancer. N Engl J Med. 2021;384(25):2394–405.
10. Litton JK, Rugo HS, Ettl J, Hurvitz SA, Gonçalves A, Lee KH, et al. Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. N Engl J Med. 2018;379(8):753–63.
11. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23(4):517–25.
12. Nguyen L, Martens JWM, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. Nat Commun. 2020;11(1):5584.
13. Ladan MM, van Gent DC, Jager A. Homologous recombination deficiency testing for brca-like tumors: The road to clinical validation. Cancers. 2021;12:1–23.
14. 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(7463):415–21.
15. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149(5):994–1007.
16. 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(7605):47–54.
17. 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(7793):94–101.
18. Macintyre G, Goranova TE, De Silva D, Ennis D, Piskorz AM, Eldridge M, et al. Copy number signatures and mutational processes in ovarian carcinoma. Nat Genet. 2018;50(9):1262–70.
19. Wang S, Li H, Song M, Tao Z, Wu T, He Z, et al. Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. PLoS Genet. 2021;17:e1009557.
20. Drews RM, Hernando B, Tarabichi M, Haase K, Lesluyes T, Smith PS, et al. A pan-cancer compendium of chromosomal instability. Nature. 2022;606(7916):976–83.
21. Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. Nature. 2022;606(7916):984–91.
22. Marquard AM, Eklund AC, Joshi T, Krzystanek M, Favero F, Wang ZC, et al. Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. Biomark Res. 2015;3(1):9.
23. 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(15):3764–73.
24. Birkbak NJ, Wang ZC, Kim JY, 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(4):366–75.
25. 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(21):5454–62.
26. 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(10):1776–82.
27. 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(5):912–9.
28. Batalini F, Gulhan DC, Mao V, Tran A, Polak M, Xiong N, et al. Mutational Signature 3 Detected from Clinical Panel Sequencing is Associated with Responses to Olaparib in Breast and Ovarian Cancers. Clin Cancer Res. 2022;28(21):4714–23.
29. Dias MP, Moser SC, Ganesan S, Jonkers J. Understanding and overcoming resistance to PARP inhibitors in cancer therapy. Nat Rev Clin Oncol. 2021;18:773–91.
30. Noordermeer SM, Adam S, Setiaputra D, Barazas M, Pettitt SJ, Ling AK, et al. The shieldin complex mediates 53BP1-dependent DNA repair. Nature. 2018;560(7716):117–21.
31. Severson TM, Wolf DM, Yau C, Peeters J, Wehkam D, Schouten PC, et al. The BRCA1ness signature is associated significantly with response to PARP inhibitor treatment versus control in the I-SPY 2 randomized neoadjuvant setting. Breast Cancer Res. 2017;19(1):99.
32. Peng G, Lin CCJ, Mo W, Dai H, Park YY, Kim SM, et al. Genome-wide transcriptome profiling of homologous recombination DNA repair. Nat Commun. 2014;20:5.
33. Daemen A, Wolf DM, Korkola JE, Griffith OL, Frankum JR, Brough R, et al. Cross-platform pathway-based analysis identifies markers of response to the PARP inhibitor olaparib. Breast Cancer Res Treat. 2012;135(2):505–17.
34. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet. 2006;38(9):1043–8.
35. Pan JW, Ng PS, Mamduh M, Zabidi A, Fatin PN, Teo JY, et al. Gene signature for predicting homologous recombination deficiency in triple-negative breast cancer. BioRxiv. 2022; Available from: https://doi.org/10.1101/2022.06.08.495296
36. Sunada S, Nakanishi A, Miki Y. Crosstalk of DNA double-strand break repair pathways in poly(ADP-ribose) polymerase inhibitor treatment of breast cancer susceptibility gene 1/2-mutated cancer. Cancer Sci. 2018;109:893–9.
37. Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harb Perspect Biol. 2015;7(4):a016600.
38. Pusztai L, Yau C, Wolf DM, Han HS, Du L, Wallace AM, et al. Durvalumab with olaparib and paclitaxel for high-risk HER2-negative stage II/III breast cancer: Results from the adaptively randomized I-SPY2 trial. Cancer Cell. 2021;39(7):989–998.e5.

39. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The International Cancer Genome Consortium Data Portal. Nat Biotechnol. 2019;37(4):367–9.
40. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44(8): e71.
41. Valieris R, Amaro L, de Toledo Osório CAB, Bueno AP, Mitrowsky RAR, Carraro DM, et al. Deep learning predicts underlying features on pathology images with therapeutic relevance for breast and gastric cancer. Cancers (Basel). 2020;12(12):1–12.
42. Kan Z, Ding Y, Kim J, Jung HH, Chung W, Lal S, et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. Nat Commun. 2018;9(1):1725.
43. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–4.
44. Wiecek AJ, Cutty SJ, Kornai D, Parreno-Centeno M, Gourmet LE, Taglia-azucchi GM, et al. Genomic hallmarks and therapeutic implications of cancer cell quiescence Running title: Hallmarks of cancer quiescence and therapeutic implications. BioRxiv. 2022; Available from: https://doi.org/10.1101/2021.11.12.468410
45. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483(7391):603–7.
46. Yu C, Mannan AM, Yvone GM, Ross KN, Zhang YL, Marton MA, et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. Nat Biotechnol. 2016;34(4):419–23.
47. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;5:8.
48. Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlioglu E, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res. 2020;30(9):745–62.
49. Bassez A, Vos H, Van Dyck L, Floris G, Arijs I, Desmedt C, et al. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. Nat Med. 2021;27(5):820–32.
50. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–3587.e29.
51. 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(1):31.
52. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. Available from: http://cran-logs.rstudio.com
53. Takaya H, Nakai H, Takamatsu S, Mandai M, Matsumura N. Homologous recombination deficiency status-based classification of high-grade serous ovarian carcinoma. Sci Rep. 2020;10(1):2757.
54. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941–7.
55. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017;171(5):1029–1041.e21.
56. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 2018;9(1):20.
57. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
58. Buffa FM, Harris AL, West CM, Miller CJ. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. Br J Cancer. 2010;102(2):428–35.
59. Bhandari V, Hoey C, Liu LY, Lalonde E, Ray J, Livingstone J, et al. Molecular landmarks of tumor hypoxia across cancer types. Nat Genet. 2019;51(2):308–18.
60. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw. 2008;28(5). Available from: http://www.jstatsoft.org/
61. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Vol. 33, JSS Journal of Statistical Software. 2010. Available from: http://www.jstatsoft.org/
62. Ulgen E, Ozisik O, Sezerman OU. PathfindR: An R package for comprehensive identification of enriched pathways in omics data through active subnetworks. Front Genet. 2019;10(SEP).
63. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(1):W90–7.
64. Rezaie N, Reese F, Mortazavi A. PyWGCNA: A Python package for weighted gene co-expression network analysis. BioRxiv. 2022; Available from: https://doi.org/10.1101/2022.08.22.504852
65. Xing X, Yang F, Li H, Zhang J, Zhao Y, Gao M, et al. Multi-level attention graph neural network based on co-expression gene modules for disease diagnosis and prognosis. Bioinformatics. 2022;38(8):2178–86.
66. Paszke A, Gross S, Massa F, Lerer A, Bradbury Google J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in neural information proceessing systems. 2019;32.
67. Fey M, Lenssen JE. Fast Graph Representation Learning with PyTorch Geometric. ArXiv. 2019 Mar 6; Available from: http://arxiv.org/abs/1903.02428
68. Efremova M, Vento-Tormo M, Teichmann SA, et al. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc 15, 1484–1506 (2020).
69. Hwang T, Reh S, Dunbayev Y, Zhong Y, Takata Y, Shen J, et al. Defining the mutation signatures of DNA polymerase θ in cancer genomes. NAR Cancer. 2020;2(3):zcaa017.
70. Perez-Villatoro F, Oikkonen J, Casado J, Chernenko A, Gulhan DC, Tumiati M, et al. Optimized detection of homologous recombination deficiency improves the prediction of clinical outcomes in cancer. NPJ Precis Oncol. 2022;6(1):96.
71. Telli ML, Stover DG, Loi S, Aparicio S, Carey LA, Domchek SM, et al. Homologous recombination deficiency and host anti-tumor immunity in triple-negative breast cancer. Breast Cancer Res Treat. 2018;171:21–31.
72. Schrempf A, Slyskova J, Loizou JI. Targeting the DNA Repair Enzyme Polymerase θ in Cancer Therapy. Trends Cancer. 2021;7:98–111.
73. Patterson-Fortin J, D'Andrea AD. Exploiting the microhomology-mediated end-joining pathway in cancer therapy. Cancer Res. 2020;80:4593–600.
74. Wiecek AJ, Cutty SJ, Kornai D, Parreno-Centeno M, Gourmet LE, Taglia-zucchi GM, et al. Genomic hallmarks and therapeutic implications of G0 cell cycle arrest in cancer. Genome Biol. 2023;24(1):128.
75. Dominguez-Sola D, Gautier J. MYC and the control of DNA replication. Cold Spring Harb Perspect Med. 2014;4(6):a014423.
76. McAlpine JN, Porter H, Köbel M, Nelson BH, Prentice LM, Kalloger SE, et al. BRCA1 and BRCA2 mutations correlate with TP53 abnormalities and presence of immune cell infiltrates in ovarian high-grade serous carcinoma. Mod Pathol. 2012;25(5):740–50.
77. Greenblatt MS, Chappuis PO, Bond JP, Hamel N, Foulkes WD. TP53 Mutations in Breast Cancer Associated with BRCA1 or BRCA2 Germ-line Mutations: Distinctive Spectrum and Structural Distribution 1. Cancer Res. 2001;61. Available from: http://metalab.unc.edu/dnam/mainpage.html.
78. Takaku M, Grimm SA, Wade PA. GATA3 in breast cancer: Tumor suppressor or oncogene? Gene Expr. 2015;16:163–8.
79. Cohen H, Ben-Hamo R, Gidoni M, Yitzhaki I, Kozol R, Zilberberg A, et al. Shift in GATA3 functions, and GATA3 mutations, control progression and clinical presentation in breast cancer. Breast Cancer Res. 2014;16(1):464.
80. Ansari-Pour N, Zheng Y, Yoshimatsu TF, Sanni A, Ajani M, Reynier JB, et al. Whole-genome analysis of Nigerian patients with breast cancer reveals ethnic-driven somatic evolution and distinct genomic subtypes. Nat Commun. 2021;12(1):6946.
81. Heijink AM, Talens F, Jae LT, van Gijn SE, Fehrmann RSN, Brummelkamp TR, et al. BRCA2 deficiency instigates cGAS-mediated inflammatory signaling and confers sensitivity to tumor necrosis factor-alpha-mediated cytotoxicity. Nat Commun. 2019;10(1):100.
82. Pawlyn C, Loehr A, Ashby C, Tytarenko R, Deshpande S, Sun J, et al. Loss of heterozygosity as a marker of homologous repair deficiency in multiple myeloma: A role for PARP inhibition? Leukemia. 2018;32(7):1561–6.

83. Gruber JJ, Afghahi A, Timms K, DeWees A, Gross W, Aushev VN, et al. A phase II study of talazoparib monotherapy in patients with wild-type BRCA1 and BRCA2 with a mutation in other homologous recombination genes. Nat Cancer. 2022;3(10):1181–91.
84. Dillon KM, Bekele RT, Sztupinszki Z, Hanlon T, Rafiei S, Szallasi Z, et al. PALB2 or BARD1 loss confers homologous recombination deficiency and PARP inhibitor sensitivity in prostate cancer. NPJ Precis Oncol. 2022;6(1):49.
85. Zhang M, Liu G, Xue F, Edwards R, Sood AK, Zhang W, et al. Copy number deletion of RAD50 as predictive marker of BRCAness and PARP inhibitor response in BRCA wild type ovarian cancer. Gynecol Oncol. 2016;141(1):57–64.
86. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017;16:495–506.
87. Miyagawa K, Tsuruga T, Kinomura A, Usui K, Katsura M, Tashiro S, et al. A role for RAD54B in homologous recombination in human cells. EMBO J. 2002;21:175–80.
88. Prakash R, Sandoval T, Morati F, Zagelbaum JA, Lim PX, White T, et al. Distinct pathways of homologous recombination controlled by the SWS1–SWSAP1–SPIDR complex. Nat Commun. 2021;12(1):4255.
89. Pearl LH, Schierz AC, Ward SE, Al-Lazikani B, Pearl FMG. Therapeutic opportunities within the DNA damage response. Nat Rev Cancer. 2015;15(3):166–80.
90. Mandal J, Mandal P, Wang TL, Shih IM. Treating ARID1A mutated cancers by harnessing synthetic lethality and DNA damage response. J Biomed Sci. 2022;29:71.
91. Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinformatics Advances. 2022;2(1):vbac016.
92. Mehibel M, Xu Y, Li CG, Moon EJ, Thakkar KN, Diep AN, et al. Eliminating hypoxic tumor cells improves response to PARP inhibitors in homologous recombination–deficient cancer models. J Clin Invest. 2021;131(11):e146256.
93. Chan N, Pires IM, Bencokova Z, Coackley C, Luoto KR, Bhogal N, et al. Contextual synthetic lethality of cancer cell kill based on the tumor microenvironment. Cancer Res. 2010;70(20):8045–54.
94. Wolff M, Kosyna FK, Dunst J, Jelkmann W, Depping R. Impact of hypoxia inducible factors on estrogen receptor expression in breast cancer cells. Arch Biochem Biophys. 2017;1(613):23–30.
95. Chu T, Wang Z, Pe'er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat Cancer. 2022;3(4):505–17.
96. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun. 2015;4:6.
97. Wang Q, Sun Z, Xia W, Sun L, Du Y, Zhang Y, et al. Role of USP13 in physiology and diseases. Front Mol Biosci. 2022;9:977122.
98. Kim W, Zhao F, Gao H, Qin S, Hou J, Deng M, et al. USP13 regulates the replication stress response by deubiquitinating TopBP1. DNA Repair (Amst). 2021;1:100.
99. Singh KK, Ayyasamy V, Owens KM, Koul MS, Vujcic M. Mutations in mitochondrial DNA polymerase-γ promote breast tumorigenesis. J Hum Genet. 2009;54(9):516–24.
100. Copeland WC. Defects of mitochondrial DNA replication. J Child Neurol. 2014;29(9):1216–24.
101. Koldobskiy MA, Chakraborty A, Werner JK, Snowman AM, Juluri KR, Scott Vandiver M, et al. p53-mediated apoptosis requires inositol hexakisphosphate kinase-2. PNAS. 2010;107(49):20947–51. https://doi.org/10.1073/pnas.1015671107.
102. Rao F, Cha J, Xu J, Xu R, Vandiver MS, Tyagi R, et al. Inositol Pyrophosphates Mediate the DNA-PK/ATM-p53 Cell Death Pathway by Regulating CK2 Phosphorylation of Tti1/Tel2. Mol Cell. 2014;54(1):119–32.
103. Vilas CK, Emery LE, Denchi EL, Miller KM. Caught with One's Zinc Fingers in the Genome Integrity Cookie Jar. Trends Genet. 2018;43:313–25.
104. Singh JK, van Attikum H. DNA double-strand break repair: Putting zinc fingers on the sore spot. Semin Cell Dev Biol. 2021;113:65–74.
105. Feeley LP, Mulligan AM, Pinnaduwage D, Bull SB, Andrulis IL. Distinguishing luminal breast cancer subtypes by Ki67, progesterone receptor or TP53 status provides prognostic information. Mod Pathol. 2014;27(4):554–61.
106. Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009;101(10):736–50.
107. Irie A, Kontani K, Kihara M, Liu D, Shirato Y, Seki M, et al. Galectin-9 as a Prognostic Factor with Antimetastatic Potential in Breast Cancer. Clin Cancer Res [Internet]. 2005;11(8):2962–8. Available from: www.aacrjournals.org
108. Bakhoum SF, Ngo B, Laughney AM, Cavallo JA, Murphy CJ, Ly P, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. Nature. 2018;553(7689):467–72.
109. Yoon HK, Kim TH, Park SG, Jung H, Quan X, Park SJ, et al. Effect of anthracycline and taxane on the expression of programmed cell death ligand-1 and galectin-9 in triple-negative breast cancer. Pathol Res Pract. 2018;214(10):1626–31.
110. Lv Y, Ma X, Ma Y, Du Y, Feng J. A new emerging target in cancer immunotherapy: Galectin-9 (LGALS9). Genes and Diseases: Chongqing University; 2022.
111. Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, et al. Genetic and transcriptional evolution alters cancer cell line drug response. Nature. 2018;560(7718):325–30.
112. Bhin J, Paes Dias M, Gogola E, Rolfs F, Piersma SR, de Bruijn R, et al. Multi-omics analysis reveals distinct non-reversion mechanisms of PARPi resistance in BRCA1- versus BRCA2-deficient mammary tumors. Cell Rep. 2023;42(5):112538.
113. França GS, Baron M, Pour M, King BR, Rao A, Misirlioglu S, et al. Drug-induced adaptation along a resistance continuum in cancer cells. BioRxiv. 2022; Available from: https://doi.org/10.1101/2022.06.21.496830
114. Funnell T, O'Flanagan CH, Williams MJ, McPherson A, McKinney S, Kabeer F, et al. Single-cell genomic variation induced by mutational processes in cancer. Nature. 2022;612(7938):106–15.
115. Budczies J, Kluck K, Beck S, Ourailidis I, Allgäuer M, Menzel M, et al. Homologous recombination deficiency is inversely correlated with microsatellite instability and identifies immunologically cold tumors in most cancer types. Journal of Pathology: Clinical Research. 2022;8(4):371–82.
116. Pellegrino B, Musolino A, Llop-Guevara A, Serra V, De Silva P, Hlavata Z, et al. Homologous Recombination Repair Deficiency and the Immune Response in Breast Cancer: A Literature Review. TransL Oncol. 2020;13:410–22.
117. Sumitani N, Ishida K, Sawada K, Kimura T, Kaneda Y, Nimura K. Identification of Malignant Cell Populations Associated with Poor Prognosis in High-Grade Serous Ovarian Cancer Using Single-Cell RNA Sequencing. Cancers (Basel). 2022;14(15):3580.
118. Launonen IM, Lyytikäinen N, Casado J, Anttila EA, Szabó A, Haltia UM, et al. Single-cell tumor-immune microenvironment of BRCA1/2 mutated high-grade serous ovarian cancer. Nat Commun. 2022;13(1):835.
119. Tietscher S, Wagner J, Anzeneder T, Langwieder C, Rees M, Sobottka B, et al. A comprehensive single-cell map of T cell exhaustion-associated immune environments in human breast cancer. Nat Commun. 2023;14(1):98. Available from: https://www.nature.com/articles/s41467-022-35238-w
120. Vikas P, Borcherding N, Chennamadhavuni A, Garje R. Therapeutic Potential of Combining PARP Inhibitor and Immunotherapy in Solid Tumors. Front Oncol. 2020;10:570.
121. Peyraud F, Italiano A. Combined parp inhibition and immune checkpoint therapy in solid tumors. Cancers. 2020;12:1–28.
122. Hong C, Schubert M, Tijhuis AE, Requesens M, Roorda M, van den Brink A, et al. cGAS–STING drives the IL-6-dependent survival of chromosomally instable cancers. Nature. 2022;607(7918):366–73.
123. Jacobson DH, Pan S, Fisher J, Secrier M. Multi-scale characterisation of homologous recombination deficiency in breast cancer. Github. https://github.com/secrierlab/MultiscaleHRD . 2022.

Publisher's Note