Published in final edited form as:

Sci Transl Med. 2014 March 26; 6(229): 229ra42. doi:10.1126/scitranslmed.3008291.

DNA Repair Pathway Gene Expression Score Correlates with Repair Proficiency and Tumor Sensitivity to Chemotherapy

Sean. P. Pitroda¹, Itai. M. Pashtan², Hillary. L. Logan¹, Brian Budke¹, Thomas E. Darga¹, Ralph. R. Weichselbaum^{1,3}, and Philip. P. Connell^{1,*}

¹Department of Radiation and Cellular Oncology, University of Chicago

²Department of Radiation Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School

³Ludwig Center for Metastasis Research, University of Chicago

Abstract

Homologous recombination (HR) and non-homologous end joining (NHEJ) are alternative pathways of double-strand DNA break repair. We developed a method to quantify the efficiency of DNA repair pathways in the context of cancer therapy. The Recombination Proficiency Score (RPS) utilizes the expression levels for four genes involved in DNA repair pathway preference (RIF1, PARI, RAD51, and Ku80), such that high expression of these genes yields a low RPS. Carcinoma cells with low RPS exhibit HR suppression and frequent DNA copy number alterations, which are characteristic of error-prone repair processes that arise in HR-deficient backgrounds. The RPS system was clinically validated in patients with breast or non-small cell lung carcinomas (NSCLC). Tumors with low RPS were associated with greater mutagenesis, adverse clinical features, and inferior patient survival rates, suggesting that HR suppression plays a central role in promoting the genomic instability that fuels malignant progression. This adverse prognosis associated with low RPS was diminished if NSCLC patients received adjuvant chemotherapy, suggesting that HR suppression and associated sensitivity to platinum-based drugs counteracts the adverse prognosis associated with low RPS. Therefore, RPS may predict which therapies will be effective for individual patients, thereby enabling more personalized oncology care.

Keywords

DNA repair; homologous recombination; genomic instability; malignant progression; personalized medicine

*To whom correspondence should be addressed. ; Email: pconnell@radonc.uchicago.edu

Author contributions: Most of the analyses were designed and performed by SPP and PPC. For characterization of the six representative cell lines, HLL measured the HR efficiencies, TED measured the mRNA levels, and BB is responsible for the determination of drug sensitivities. The manuscript was drafted by PPC and revised by SPP, IMP, and RRW.

Competing interests: None of the authors have conflicts to report.

Data and materials availability: N/A

Introduction

Homologous recombination (HR) and non-homologous end-joining (NHEJ) are competing pathways that repair double-stranded DNA breaks (DSBs) generated by certain cancer treatment modalities. HR also serves additional functions such as promoting cellular tolerance to DNA-damaging drugs that disrupt replication forks⁽¹⁾. Both HR and NHEJ facilitate DNA repair following the recruitment of upstream sensor/effector proteins (Figure 1). The HR pathway catalyzes DSB repair by identifying a stretch of homologous DNA and by replicating from this homologous DNA template, while NHEJ repairs DSBs by processing and re-ligating the DSB ends^(1, 2).

When faced with a DSB, the cell's decision of whether to utilize HR vs. NHEJ is influenced by the cell cycle stage. NHEJ is the dominant pathway for repairing DSBs during G0/G1 stages of the cell cycle, while HR occurs generally during S and G2. This regulation of repair is governed primarily by BRCA1 and 53BP1 proteins, which compete for occupancy at the DSB site⁽³⁾. Stabilization of 53BP1 in cooperation with Rif1 leads to the exclusion of BRCA1 protein from the repair complex, and the DSB subsequently progresses to repair by NHEJ^(4, 5). If 53BP1 is excluded from the repair complex, then the DSB progresses to repair by HR. In this case, the DSB ends are processed into HR substrates, which involves 5' to 3' nuclease activity that generates 3' single-stranded DNA tails. This end processing is promoted by several proteins including CtIP, BRCA1, and the MRN (Mre11/RAD50/NBS1) complex. The nuclease activity is also specifically triggered by interactions between Mre11 and cyclin dependent kinase 2, thereby promoting the phosphorylation of CtIP preferentially in S/G2 cells⁽⁶⁾.

The efficiencies of these repair processes have important implications for carcinogenesis and malignant tumor progression. Like HR, the canonical version of NHEJ is thought to repair DNA with high fidelity^(7, 8). However, some DSBs can undergo extensive degradation prior to re-ligation using processes termed microhomology-mediated end joining and single-strand annealing, both of which create mutagenic deletions^(8, 9). Similarly, mutations can arise if replication-disrupting lesions are not properly repaired prior to DNA replication, in which case these lesions may prompt homology-mediated polymerase template switching⁽¹⁰⁾. So tumors that harbor ineffective error-free DNA repair machinery are likely to exhibit greater genomic instability, which is expected to drive malignant progression and generate more aggressive tumor phenotypes. A method that predicts error-free repair proficiency from human tumor biopsy tissues might have broad applications in clinical oncology as a prognostic indicator, since genetic instability may indicate a greater propensity for malignant phenotypes like metastagenicity.

The cellular efficiencies of these repair processes can also directly impact tumor responsiveness during the treatment of cancer patients. The most striking examples are the hypersensitivities of HR-deficient tumors to PARP inhibitors⁽¹¹⁻¹³⁾ or platinum-based chemotherapies^(14, 15). At present, methods to measure HR or NHEJ proficiency from human tumor biopsy tissues are limited^(16, 17). Some studies have measured the rate of DSB rejoining in tumors (e.g. H2AX phosphorylation kinetics), and rapid DSB rejoining may predict resistance of human tumors to radiotherapy and some chemotherapy drugs

(reviewed in (¹⁸)). However, a method that successfully predicts repair efficiency might have important applications in clinical oncology, since it would predict sensitivity of tumors to specific classes of treatment.

Human tumors exhibit a wide range of malignant features and responsiveness to treatments that damage DNA. We hypothesized that a component of this variability can be explained by differential efficiencies of DNA repair pathways. To study this further we developed an analytic tool to indirectly quantify the efficiency of HR in individual cancers. This scoring system relies on the expression of four DNA repair genes: Rif1, PARI, RAD51, and Ku80. We show here that the Recombination Proficiency Score (RPS) correctly predicts sensitivity to specific classes of chemotherapy, correlates with degree of genomic instability within tumor cells, and provides valuable prognostic and predictive information that is not available using existing diagnostic methods.

Results

Development of the RPS system

We sought to create a method that predicts the efficiency of HR repair within any given cancer. To accomplish this, we developed a scoring system that correlates gene expression patterns with HR proficiency in human cancer cell lines. Gene expression levels and corresponding drug sensitivity data were collected from the Broad-Novartis Cancer Cell Line Encyclopedia (CCLE) (¹⁹). Given the wide biological diversity known to exist between different classes of human malignancies, we limited this analysis to cell lines derived only from carcinomas. Cellular resistance to the topoisomerase-I inhibiting drug topotecan was selected as a surrogate marker for HR proficiency. Topotecan is a derivative of camptothecin, and this class of drugs was selected because it disrupts replication forks and exerts toxicity preferentially in cells that harbor HR defects(^{20, 21}). Topotecan sensitivity data were available for 279 of the 634 carcinoma cell lines.

To focus our analysis on the primary cellular features that mediate specific phenotypes, we restricted the analysis to genes with direct relevance to replication stress and the DSB repair pathways (Figure 1). We further limited the analysis to 33 central proteins that participate in cellular preference toward HR vs. NHEJ, following the ataxia telangiectasia mutated (ATM) and/or ataxia telangiectasia and Rad3-related protein (ATR) activation steps of DNA damage response. Levels of mRNA were available for all of these genes except Ku70. Secondary regulators of damage response (like TP53, PTEN, and cell cycle checkpoint genes) were not considered as gene candidates for the scoring system, since they exert cellular influences that extend beyond the scope of replication stress and DSB repair. Pearson's correlation analyses demonstrated that 12 of the final list of 32 candidate genes had expression levels that significantly correlated (defined as $p < 0.05$) with cellular sensitivity to topotecan (Supplemental Table 1). In all 12 cases, increasing gene expression levels directly correlated with increasing topotecan sensitivity.

HR-related genes are highly expressed in cancer cells that harbor low HR efficiency

Rif1, Ku80, and PARI were among the genes whose expression most strongly associated with topotecan sensitivity. Rif1 and Ku80 are known to promote NHEJ and antagonize HR^(4, 5, 9). PARI is a helicase capable of disrupting RAD51 nucleofilaments, and it has been reported to antagonize HR repair⁽²²⁾.

Topotecan sensitivity also correlated with the overexpression of a family of HR-related genes, including RAD51, BLM, BRCA1, RAD51-AP1, RAD54B, PLK1, BRCA2, RAD51C, and PALB2. This observation appears surprising on the surface, since RAD51 and many of these RAD51-associated proteins are generally considered to promote HR. However, RAD51 overexpression has been previously shown to occur in the setting of HR defects caused by BRCA mutations^(23, 24). To investigate a possible connection between BRCA mutation phenotypes and our observed expression patterns, we analyzed gene expression levels in CCLE cell lines that harbor BRCA1 (HCC1937 and MDA-MB436) or BRCA2 (CAPAN1) mutations. Consistent with published observations in BRCA1-mutant human tumors⁽²³⁾, these three cell lines significantly overexpressed RAD51 and RAD51AP1. In addition, we found that BRCA-defective cells significantly overexpress additional genes known to promote various mechanisms required for HR, including CtIP which promotes the 5' to 3' ssDNA end resection⁽²⁵⁾, Plk1 which promotes the phosphorylation of 53BP1 and RAD51^(26, 27), and several genes (XRCC2, XRCC3, PALB2) that promote RAD51 filament assembly⁽²⁸⁾. These data suggest that BRCA-defective cells respond to their HR defects by increasing the expression of a fairly broad array of HR-related genes. The overexpression of HR genes as a compensatory mechanism has been proposed previously, particularly since RAD51 overexpression is known to partially suppress the HR defects that occur when key HR genes are mutated^(23, 29).

These findings were used to refine the list of genes to be used in the Recombination Proficiency Score (RPS). We hypothesized that when HR-deficiency occurs in wild type BRCA backgrounds, cells respond via compensatory overexpression of HR-related genes that mirrors the phenotypes observed in BRCA mutant cells. As such, we reasoned that many of the HR-related genes were reporting redundant predictive information in response to low HR proficiency. Gene expression levels were combined to generate a single model that predicts topotecan sensitivity, starting with genes that have known HR-antagonizing activities (Rif1, Ku80, and PARI) in order of their independent predictive power (Supplemental Table 2). The family of HR-related genes was then subsequently added incrementally into this model. The addition of RAD51 improved the model's correlation with topotecan sensitivity (relative to the initial 3 genes), however the inclusion of additional HR-related genes did not further improve the correlation. Therefore, the final four genes selected to derive the RPS were Rif1, PARI, Ku80, and RAD51.

Elevated mRNA levels for any of these genes correlated with greater sensitivity to topotecan. The RPS was defined as the sum of these four expression levels multiplied times -1, using the log2 transformed mRNA values of each gene normalized to the median mRNA within the starting 634 carcinoma cell lines. The median RPS score within the carcinoma cell lines was approximately zero, the bottom 25th percentile of RPS scores were less than -1.08, and the top 25th percentile of RPS scores were greater than 1.2.

Interestingly, CCLE cell lines with low RPS scores did indeed overexpress a broad array of HR-related genes (Figure 2). These data support the existence of a compensatory mechanism that responds to low HR efficiency. Furthermore, these results suggest that this compensatory mechanism is not limited to only the most extreme HR defects, like those resulting from BRCA mutations.

RPS predicts HR proficiency in individual cancer cell lines

The predictive value of the RPS was further tested based on sensitivity to different types of chemotherapeutic agents. Similar to results with topotecan, low RPS scores correlated to sensitivity to irinotecan, another topoisomerase-I inhibiting drug (Figure 3A). This is expected, since topoisomerase-I inhibitors generate replication fork disruptions, which require HR for repair^(20, 21). As a control, this analysis was repeated using the non-DNA damaging drug paclitaxel, and RPS did not show a correlation with sensitivity to this agent. These results support the specificity of RPS to DNA-related damage and repair. It should be noted that complete drug sensitivity data was not available for all three chemotherapy agents in all cell lines evaluated (see Supplemental Figure 1 for breakdown). However, comparable results were observed when the analyses were repeated on the subset of 137 cell lines that were tested with all three agents.

The ability of RPS to predict repair pathway preference was further tested by measuring HR repair efficiency in representative cell lines with low RPS (RKO, DU 145, COLO 205) or with mid/high RPS (PC3, HCC44, NCI-H650). These cell lines exhibited expected levels of sensitivity to topotecan and paclitaxel when independently re-tested in our laboratory (Supplemental Figure 2), which were comparable to the sensitivities mined from the CCLE database. These six cell lines were tested using a modified version of the previously described DR-GFP reporter method⁽³⁰⁾. This method utilizes a reporter DNA construct that carries two non-functional copies of green fluorescence protein (GFP), one of which is interrupted by an I-SceI endonuclease site. Induction of a DSB at the I-SceI site can lead to repair by homologous gene conversion that generates a functional copy of GFP. As demonstrated in Figure 3B, RPS correlated with HR efficiency on linear regression analysis ($R^2=0.833$, two-sided $p=0.003$). For consistency with the other results, RPS values for these cells were calculated using array-based mRNA levels from the CCLE database. We verified the identity of our six cell lines by short tandem repeat profiling (Genetic Resources Core Facility at Johns Hopkins School of Medicine), and independent quantitation by real time qRT-PCR generated mRNA measurements that were comparable to the mRNA levels reported in the CCLE database (Supplemental Figure 3).

Cell lines with high RPS have elevated genomic instability

HR plays a central role in maintaining genomic stability in cells. We hypothesized that cells with low RPS would exhibit more genome instability than cells with high RPS. To test this hypothesis, SNP array-based DNA copy number variations (CNVs) were analyzed using CCLE carcinoma cell lines (Figure 4). Low RPS scores were associated with more frequent DNA amplifications. This finding is consistent with published analyses of HR-defective cell lines, showing that mutations in RAD51D or XRCC3 promote DNA amplifications⁽³¹⁾. These amplifications are proposed to result from stress-induced replication fork disruption

and subsequent homology-mediated polymerase template switching^(7, 10). A study in RAD51 defective *S. cerevisiae* demonstrated that cells with deregulated HR frequently channel DSBs into repair by non-allelic break-induced replication, thereby stimulating the formation of segmental duplications⁽³²⁾. Additionally, we found that cells with low RPS harbored relatively frequent DNA deletions. Deletions are characteristic of error prone repair processes like microhomology-mediated end joining and single-strand annealing^(8, 9). Of note, the distributions of CNV sizes were not strongly influenced by RPS. Taken together, these results suggest that low RPS cells have reduced HR proficiency and rely more on error-prone processes to rejoin DSBs and/or to tolerate replication stress.

Mutations in TP53 are also known to exert major influences on cellular resistance to DNA damaging therapies and genomic instability. Additionally, TP53 mutation status has been shown to influence HR efficiency^(33, 34). Therefore, we re-examined RPS -associated CNVs in the context of TP53 mutation status. The average RPS was not significantly different between the 238 TP53 WT cell lines and the 386 TP53 mutant cell lines (0.25 vs. 0.41, $p=0.41$). Also, the association between increased CNVs and low RPS was observed in both TP53 WT and mutant cell line groups. The magnitude of RPS dependence was less pronounced in TP53 mutant cells, due to a high background of deletions in TP53 mutant cells. A high deletion frequency is not surprising in TP53 mutant cells, since deletions are known to occur 40-300 times more following TP53 inactivation⁽³⁵⁾. These data suggest, therefore, that TP53 mutation status and RPS offer independent predictive information regarding genomic instability.

A possible relationship between RPS and TP53 status was further studied by examining resistance to DNA damage. The ability of RPS to associate with topotecan sensitivity on logistic regression was similar in both TP53 WT and mutant cell line subgroups ($p<0.003$ for both). This association supports the role of RPS as a predictor of HR proficiency, which is distinct from TP53-dependent activities like apoptotic threshold modulation and cell cycle regulation.

Human tumors with low RPS exhibit unfavorable clinical characteristics and elevated genomic instability

The RPS system was clinically validated using tumor datasets from the Cancer Genome Atlas (CGA). Breast and non-small cell lung cancer (NSCLC) tumor types were selected for this analysis, because these datasets contained large sample sizes, annotations of clinical features, SNP array-based DNA CNV data, and adequate details on patient outcomes. Although some differences existed between different cancer types, tumors with lower RPS generally exhibited adverse clinical characteristics (Table 1). Low RPS tumors tended to be more locally/regionally advanced and to harbor more frequent TP53 mutations. For example, the lower quartile RPS tumors were significantly more likely to have lymph node invasion in non-small cell lung cancers ($p=0.008$). Similarly, breast cancers with low RPS commonly exhibited estrogen receptor loss ($p=0.0001$) and HER2 amplification ($p=0.007$).

These adverse features associated with low RPS may be the result of low-fidelity repair processes, which in turn promote genomic instability and malignant progression. To explore this hypothesis we analyzed CNV as a function of RPS using these same two CGA tumor

datasets. Both carcinoma types exhibited at least one class of elevated CNV in the setting of low RPS (Figure 5). This RPS -associated genome instability was observed in both TP53 WT and mutant tumors. These results suggest that mutagenic DNA repair processes dominate in low RPS tumors, thereby promoting the evolution of malignant clinical features.

RPS is prognostic and predictive of treatment sensitivity in clinical tumors

Next we evaluated whether RPS can predict clinical outcomes in human tumors. NSCLC was considered an appealing tumor type for this analysis, since NSCLC-directed chemotherapy regimens are generally platinum-based and since lung cancer is a leading cause of cancer mortality. We also sought to distinguish the prognostic and predictive utilities of RPS. Specifically, we hypothesized that low RPS would confer a poor prognosis, because of elevated mutagenesis and associated adverse tumor features. However, we also hypothesized that sensitivity to platinum-based chemotherapeutic agents is expected to simultaneously render low RPS tumors treatment-sensitive, given that HR-defective cells are hypersensitive to DNA cross-linkers. These opposing effects were predicted to counteract one another in low RPS tumors treated with chemotherapy.

The power of RPS to predict outcome in NSCLC patients was investigated using data from the JBR.10 clinical trial, which had previously demonstrated a benefit to adjuvant chemotherapy in early-stage NSCLC⁽³⁶⁾. Specifically, JBR.10 had randomly assigned patients to receive cisplatin + vinorelbine chemotherapies vs. no further treatment, following the resection of stage I-II NSCLCs. This dataset was ideal for our analysis because of its prospective randomized trial design, combined with uniform treatment details. As such, it does not suffer from the biases intrinsic to retrospectively collected datasets. In patients whose treatment consisted of surgery only, low RPS predicted for inferior 5-year overall survival relative to higher RPS (15% vs. 60%, $p = 0.004$, log rank test). This clinically validates the prognostic power of RPS (Figure 6A). Chemotherapy significantly improved 5-year overall survival in low RPS tumors (15% vs. 77%, $p = 0.01$) but did not in high RPS tumors (60% vs. 72%, $p = 0.55$). This clinically validates the ability of RPS to predict sensitivity to platinum-based chemotherapy.

These data suggest that the poor prognoses associated with low RPS might be negated by chemotherapy, since low RPS tumors are especially sensitivity to platinum-based chemotherapy. In the JBR.10 trial, for example, patients treated with chemotherapy had similar 5-year overall survival rates regardless of low vs. higher RPS (77% vs. 72%, $p = 0.70$). To study this further, we selected three additional datasets containing retrospectively collected data on NSCLC patients^(37,39). After controlling for stage on multivariate analysis, low RPS was again associated with poor survival in patients treated with surgery alone (Figure 6B). Specifically, we combined data from all four datasets using previously described methodology⁽⁴⁰⁾ and found that low RPS confers a continuous hazard ratio of 1.24 (95% CI = 1.12- 1.36). When this analysis was repeated on patients treated with surgery plus adjuvant chemotherapy, the poor prognosis associated with low RPS was diminished (hazard ratio = 0.94, 95% CI = 0.69-1.21). Taken together, these findings support the hypothesis that patients with low RPS tumors have adverse underlying prognoses, but that HR suppression and associated sensitivity to platinum-based drugs counteracts these

adverse prognostic features. Therefore, RPS predicts which therapies will be effective for individual patients, thereby enabling more personalized oncology care.

Discussion

The RPS is a novel scoring system that quantifies the expression of four genes to predict DSB repair pathway preference. Low RPS can identify tumors that harbor HR suppression and hypersensitivity to specific classes of chemotherapeutic agents. Since this scoring system provides individualized predictive information, the RPS could potentially be used to guide which classes of oncologic treatment are best suited for individual patients. The strategy used to develop this system is fundamentally different from the larger genomic characterizations of human cancers, which commonly catalogue the molecular features of particular cancer types⁽⁴¹⁾. The RPS also differs from gene expression signatures that derive predictive data from un-biased genome-wide data, in that we focused on a limited set of genes with known relevance to particular DNA repair pathways that enabled hypothesis-based analyses. An additional important difference of our study is that the predictive information provided by the RPS applies to a broad range of carcinoma types, and it may provide similar information in other non-carcinoma malignancies as well.

Looking beyond the therapeutic implications of RPS, our results have important inferences to the basic biology of malignant tumor progression. Specifically, tumors with low RPS exhibit greater mutagenesis and adverse patient outcomes. These findings elucidate a potential pathway of carcinogenesis, in which the repression of HR efficiency fuels the evolution of genomic instability and malignant progression. This is consistent with prior studies that have observed higher expression of some DNA repair genes (including RAD51 and Ku80) in metastases compared to primary tumors^(42, 43). This concept is also similar to the cancer-prone phenotypes seen in BRCA-inactivating mutations, as well as the BRCA-like phenotypes reported to occur in specific tumors like triple-negative breast cancers. A key difference, however, is that BRCA-related cancer diagnoses are relatively uncommon and distinct entities. By contrast, RPS -associated mutagenesis is a continuous effect, whereby mutagenesis gradually rises as RPS values fall. Furthermore, RPS -associated mutagenesis pertains to a broad range of cancer types. Taken together, these findings imply that HR suppression may play a common and central role in cancer development and malignant progression of tumors.

The overexpression of Rif1, Ku80, and PARI are all expected to antagonize HR^(4, 5, 9, 22). The more challenging observation, however, is the association between low HR efficiency and RAD51 overexpression. Our results suggest that cancer cells respond to HR defects by increasing the expression of a fairly broad array of HR-related genes. This concept of compensatory gene expression is not new, and RAD51 overexpression has been previously shown to functionally compensate for HR defects^(23, 29). However, the large number of HR-related genes that are overexpressed in HR-deficient cells is a new observation that builds on current knowledge. It suggests the existence of a coordinated gene expression mechanism, which extends beyond the known promoter elements that are shared by BRCA and RAD51 genes⁽⁴⁴⁾. A recent report showed that MEN1 protein can modulate HR by stimulating the transcription of several HR genes, including BRCA1, RAD51, and

RAD51AP1⁽⁴⁵⁾. However MEN1 activities are unlikely to explain our results, since MEN1 mRNA levels did not significantly correlate with RPS in CCLE cell lines (data not shown).

Alternative processes may also explain the association between HR protein overexpression and HR suppression/mutagenesis. RAD51 overexpression might be a direct cause of the HR suppression. Several studies have demonstrated that cells exhibit lower HR efficiency and reduced viability when RAD51 is experimentally overexpressed to very high levels^(23, 46, 47). Furthermore, human cancer cell lines that overexpress RAD51 to very high levels can exhibit nuclear foci of RAD51 in the absence of exogenous DNA damage⁽⁴⁸⁾, and these structures are thought to represent toxic RAD51 aggregates on undamaged chromatin that can lead to genomic instability⁽⁴⁹⁾. High RAD51 levels may also contribute to genome instability by catalyzing the formation of DNA-RNA hybrids, whereby an RNA transcript invades a dsDNA helix⁽⁵⁰⁾. Another possibility is that overexpressed HR proteins might promote mutagenesis by enabling homology-driven error-prone processes, like single-strand annealing, replication template switching, and non-allelic HR. For example, transient overexpression of RAD51 was shown to promote the formation of aberrant homology-mediated repair products, involving gene conversion events that lead to chromosomal translocations⁽⁵¹⁾. Likewise, RAD52 protein catalyzes the annealing of homology-containing oligonucleotides biochemically⁽⁵²⁾, and RAD52 has been shown to promote DSB rejoining via single-strand annealing in human cells⁽⁹⁾. In another example, the HR-promoting protein RAD51AP1 was found to be a key component of an expression signature that predicts for chromosomal aberrations⁽⁵³⁾. Therefore, HR protein overexpression may directly stimulate homology-mediated events that contribute to mutagenesis in low RPS tumor cells.

The predictive ability of this four gene system was more powerful than larger combinations of DNA repair gene expression values that we tested. While the simplicity of the RPS is appealing, the mechanisms governing DSB repair pathways are complex. Indeed the four RPS -defining proteins are major determinants of pathway choice. However, there is an ever-growing list of other proteins that play important roles in pathway choice, including BRCA1, 53PB1, and CtIP. Furthermore, at least some of these DNA damage response proteins can undergo damage-induced post-translational modifications by sumoylation, ubiquitylation, phosphorylation, and proteasome-mediated degradation^(6, 54, 55). These processes cannot be effectively measured with mRNA expression levels alone, as was used to calculate RPS. Despite these potential limitations, however, the four mRNA-based RPS system does successfully provide powerful predictive data.

In conclusion, these results have broad implications for cancer biology and oncologic patient care. These results suggest that HR suppression plays a common and central role in malignant tumor progression. Second, RPS can predict tumor sensitivity to different classes of therapy, so this system may enable a transition toward personalized oncology care. RPS can identify tumors that harbor HR suppression and hypersensitivity to certain chemotherapeutic classes, like inter-strand DNA cross-linkers, topoisomerase inhibitors, and PARP inhibitors. Finally, RPS might be utilized to select appropriate candidates for other investigational drugs that directly target individual components of the DNA repair machinery (such drugs are reviewed in ⁽⁵⁶⁾).

Materials and Methods

Development of RPS scoring system in human cancer cell lines and association with drug sensitivity

Data for mRNA expression, copy number variation, and drug sensitivity for human carcinoma cell lines (n=634) were collected from the Broad-Novartis Cancer Cell Line Encyclopedia (CCLE). Robust Multi-array Average (RMA)-normalized mRNA expression values were normalized to the median value across all carcinoma samples and subsequently log2 transformed. SNP array-based DNA copy number values were filtered to eliminate individual SNPs. For CNV analysis, minimum deletion size was defined as copy number segment mean -0.6 , while minimum insertion size was defined as copy number segment mean $+1.4$ ($\log_2$ [copy number/2]). Deletions and insertions were binned by size, whereby bins represent 10-fold increments in size. Drug sensitivities for topotecan (n=279), irinotecan (n=180), and paclitaxel (n=257) were determined by IC₅₀ values. IC₅₀ values ≥ 8 μM were outliers and, therefore, censored from the analysis. TP53 mutation status was determined by hybrid capture sequencing data, which was available for all carcinoma cell lines. In the six cell lines used for HR reporter experiments (RKO, DU 145, COLO 205, PC3, HCC44, NCI-H650), sensitivities to topotecan and paclitaxel were confirmed in our lab using an acute continuous 3-day exposure of cells to drugs; this method is identical to the method that was used to generate the CCLE drug sensitivity data.

Quantification of HR efficiency in cells

An HR reporter-containing plasmid (pDR-GFP), an I-Sce I expressing plasmid (pC β ASce), and an empty vector control plasmid (pCAG) were provided by Maria Jasin. Cells transiently co-transfected with combinations of either pDR-GFP + pC β ASce or pDR-GFP + pCAG. To accomplish this, 0.5×10^6 cells at 80% confluence were electroporated with 15 μg of each plasmid in 4mm cuvettes, using the following settings: 325-375 V, 975 μF . Electroporation voltages were optimized in order to minimize differences in transfection efficiencies between the six cell lines. Cells were transferred into the appropriate complete growth medium and allowed to grow for 48 hours, following which they were analyzed with a Becton-Dickinson FACSscan. Live cells were collected based on size/complexity and 7-aminoactinomycin D (7-AAD) exclusion. The fraction of live cells exhibiting GFP positivity was quantified. To account for any remaining differences that persisted in transfection efficiencies between cell lines, the GFP positivity resulting from pDR-GFP + pC β ASce transfection was normalized to GFP positivity resulting from pDR-GFP + pCAG transfection. Experiments were performed in triplicate, and the displayed error bars denote standard error.

Evaluation of RPS in human tumor datasets and association with clinical characteristics

Breast and non-small cell lung cancer (NSCLC) tumor datasets were collected from the Cancer Genome Atlas (CGA). Stage IV patients with metastatic disease or those patients without a specified stage were excluded from analysis. TP53 mutation status was determined by SNP array-based DNA copy number data. Normalized mRNA expression, copy number variation, and TP53 mutation status were available for 295 breast cancers and 153 NSCLCs. CNV analysis was performed as described for the CCLE carcinoma cell lines. Clinical

characteristics and prognostic factors were available for 280 breast cancers and 145 NSCLCs with available mRNA expression data.

Quantification of prognostic and predictive value of RPS

Four publicly available NSCLC datasets were collected from Gene Expression Omnibus (GEO; accession numbers GSE14814 [JBR.10 trial], GSE31210 [Japanese National Cancer Center Research Institute], and GSE42127 [MD Anderson Cancer Center]) and from the National Cancer Institute caArray website at <https://array.nci.nih.gov/caarray/project/jacob-00182> [Director's Challenge Consortium]. mRNA expression values were normalized to the median value across all patient samples within each respective dataset and subsequently log2 transformed. Patient samples were grouped based on type of treatment. In total, 581 patients underwent surgery alone and 164 patients received surgery + chemotherapy. Cox proportional hazard analysis for overall survival was used to determine the hazard ratio for the RPS as a continuous variable. All NSCLC dataset analyses were limited to stage I and II patients.

Statistical analysis

All analyses were performed using JMP 9.0 (SAS Institute Inc.; Cary, NC). A p-value 0.05 was considered statistically significant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are deeply indebted to the curators of publically available datasets, including the CGA, CCLE, and NSCLC tumor datasets. We thank Douglas Bishop and Jennifer Mason for critical reviews of the manuscript.

Funding: This work was supported by funding from the National Institutes of Health [CA142642-02 2010-2015 to PPC], Ludwig Foundation for Cancer Research [RRW], and the Lung Cancer Research Foundation [RRW].

References

1. Thompson LH, Schild D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. *Mutat Res.* 2001; 477:131. [PubMed: 11376695]
2. Lieber MR, Ma Y, Pannicke U, Schwarz K. The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination. *DNA Repair (Amst).* Aug-Sep;2004 3:817. [PubMed: 15279766]
3. Chapman JR, Sossick AJ, Boulton SJ, Jackson SP. BRCA1-associated exclusion of 53BP1 from DNA damage sites underlies temporal control of DNA repair. *J Cell Sci.* Aug 1.2013 125:3529. [PubMed: 22553214]
4. Zimmermann M, Lottersberger F, Buonomo SB, Sfeir A, de Lange T. 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science.* Feb 8.2013 339:700. [PubMed: 23306437]
5. Chapman JR, et al. RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Mol Cell.* Mar 7.2013 49:858. [PubMed: 23333305]
6. Buis J, Stoneham T, Spehalski E, Ferguson DO. Mre11 regulates CtIP-dependent double-strand break repair by interaction with CDK2. *Nat Struct Mol Biol.* Feb.2012 19:246. [PubMed: 22231403]

7. Arlt MF, Rajendran S, Birkeland SR, Wilson TE, Glover TW. De novo CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining. *PLoS Genet.* Sep.2012 8:e1002981. [PubMed: 23028374]
8. Guirouilh-Barbat J, et al. Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. *Mol Cell.* Jun 4.2004 14:611. [PubMed: 15175156]
9. Bennardo N, Cheng A, Huang N, Stark JM. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet.* Jun.2008 4:e1000110. [PubMed: 18584027]
10. Malkova A, Haber JE. Mutations arising during repair of chromosome breaks. *Annual review of genetics.* 2012; 46:455.
11. Bryant HE, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature.* Apr 14.2005 434:913. [PubMed: 15829966]
12. Farmer H, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature.* Apr 14.2005 434:917. [PubMed: 15829967]
13. O'Shaughnessy J, et al. Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *N Engl J Med.* Jan 20.2011 364:205. [PubMed: 21208101]
14. Edwards SL, et al. Resistance to therapy caused by intragenic deletion in BRCA2. *Nature.* Feb 28.2008 451:1111. [PubMed: 18264088]
15. Sakai W, et al. Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature.* Feb 28.2008 451:1116. [PubMed: 18264087]
16. Willers H, et al. Utility of DNA repair protein foci for the detection of putative BRCA1 pathway defects in breast cancer biopsies. *Mol Cancer Res.* Aug.2009 7:1304. [PubMed: 19671671]
17. Birkelbach M, et al. Detection of impaired homologous recombination repair in NSCLC cells and tissues. *J Thorac Oncol.* Mar.2013 8:279. [PubMed: 23399959]
18. Redon CE, et al. gamma-H2AX and other histone post-translational modifications in the clinic. *Biochim Biophys Acta.* Jul.2012 1819:743. [PubMed: 22430255]
19. Barretina J, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* Mar 29.2012 483:603. [PubMed: 22460905]
20. Nitiss J, Wang JC. DNA topoisomerase-targeting antitumor drugs can be studied in yeast. *Proc Natl Acad Sci U S A.* Oct.1988 85:7501. [PubMed: 2845409]
21. Arnaudeau C, Lundin C, Helleday T. DNA double-strand breaks associated with replication forks are predominantly repaired by homologous recombination involving an exchange mechanism in mammalian cells. *J Mol Biol.* Apr 13.2001 307:1235. [PubMed: 11292338]
22. Moldovan GL, et al. Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Mol Cell.* Jan 13.2012 45:75. [PubMed: 22153967]
23. Martin RW, et al. RAD51 up-regulation bypasses BRCA1 function and is a common feature of BRCA1-deficient breast tumors. *Cancer Res.* Oct 15.2007 67:9658. [PubMed: 17942895]
24. Honrado E, et al. Immunohistochemical expression of DNA repair proteins in familial breast cancer differentiate BRCA2-associated tumors. *J Clin Oncol.* Oct 20.2005 23:7503. [PubMed: 16234517]
25. Sartori AA, et al. Human CtIP promotes DNA end resection. *Nature.* Nov 22.2007 450:509. [PubMed: 17965729]
26. Yata K, et al. Plk1 and CK2 act in concert to regulate Rad51 during DNA double strand break repair. *Mol Cell.* Feb 10.2012 45:371. [PubMed: 22325354]
27. van Vugt MA, et al. A mitotic phosphorylation feedback network connects Cdk1, Plk1, 53BP1, and Chk2 to inactivate the G(2)/M DNA damage checkpoint. *PLoS Biol.* Jan.2010 8:e1000287. [PubMed: 20126263]
28. Thompson LH. Recognition signaling and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat Res.* Oct-Dec;2012 751:158. [PubMed: 22743550]
29. Takata M, et al. Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. *Mol Cell Biol.* 2001; 21:2858. [PubMed: 11283264]
30. Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* 1999; 13:2633. [PubMed: 10541549]

31. Hinz JM, et al. Repression of mutagenesis by Rad51D-mediated homologous recombination. *Nucleic Acids Res.* 2006; 34:1358. [PubMed: 16522646]
32. Payen C, Koszul R, Dujon B, Fischer G. Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet.* 2008; 4:e1000175. [PubMed: 18773114]
33. Linke SP, et al. p53 interacts with hRAD51 and hRAD54, and directly modulates homologous recombination. *Cancer Res.* May 15.2003 63:2596. [PubMed: 12750285]
34. Sirbu BM, et al. ATR-p53 restricts homologous recombination in response to replicative stress but does not limit DNA interstrand crosslink repair in lung cancer cells. *PLoS One.* 2011; 6:e23053. [PubMed: 21857991]
35. Gebow D, Miselis N, Liber HL. Homologous and nonhomologous recombination resulting in deletion: effects of p53 status, microhomology, and repetitive DNA length and orientation. *Mol Cell Biol.* Jun.2000 20:4028. [PubMed: 10805745]
36. Zhu CQ, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol.* Oct 10.2010 28:4417. [PubMed: 20823422]
37. Okayama H, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* Jan 1.2012 72:100. [PubMed: 22080568]
38. Tang H, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res.* Mar 15.2013 19:1577. [PubMed: 23357979]
39. Shedden K, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med.* Aug.2008 14:822. [PubMed: 18641660]
40. Neyeloff JL, Fuchs SC, Moreira LB. Meta-analyses and Forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis. *BMC Res Notes.* 2012; 5:52. [PubMed: 22264277]
41. Steensma DP. The Beginning of the End of the Beginning in Cancer Genomics. *N Engl J Med.* May 1.2012
42. Winpenninckx V, et al. Gene expression profiling of primary cutaneous melanoma and clinical outcome. *J Natl Cancer Inst.* Apr 5.2006 98:472. [PubMed: 16595783]
43. Sarasin A, Kauffmann A. Overexpression of DNA repair genes is associated with metastasis: a new hypothesis. *Mutat Res.* Jul-Aug;2008 659:49. [PubMed: 18308619]
44. Bindra RS, Glazer PM. Repression of RAD51 gene expression by E2F4/p130 complexes in hypoxia. *Oncogene.* Mar 29.2007 26:2048. [PubMed: 17001309]
45. Fang M, et al. MEN1 is a Melanoma Tumor Suppressor that Preserves Genomic Integrity by Stimulating Transcription of Genes that Promote Homologous Recombination-Directed DNA Repair. *Mol Cell Biol.* May 6.2013
46. Kim PM, Allen C, Wagener BM, Shen Z, Nickoloff JA. Overexpression of human RAD51 and RAD52 reduces double-strand break-induced homologous recombination in mammalian cells. *Nucleic Acids Res.* Nov 1.2001 29:4352. [PubMed: 11691922]
47. Flygare J, et al. Effects of HsRad51 overexpression on cell proliferation cell cycle progression and apoptosis. *Exp Cell Res.* Aug 1.2001 268:61. [PubMed: 11461118]
48. Raderschall E, et al. Elevated levels of Rad51 recombination protein in tumor cells. *Cancer Res.* Jan 1.2002 62:219. [PubMed: 11782381]
49. Shah PP, et al. Swi2/Snf2-related translocases prevent accumulation of toxic Rad51 complexes during mitotic growth. *Mol Cell.* Sep 24.2010 39:862. [PubMed: 20864034]
50. Wahba L, Gore SK, Koshland D. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *Elife.* 2013; 2:e00505. [PubMed: 23795288]
51. Richardson C, Stark JM, Ommundsen M, Jasin M. Rad51 overexpression promotes alternative double-strand break repair pathways and genome instability. *Oncogene.* Jan 15.2004 23:546. [PubMed: 14724582]
52. Mortensen UH, Bendixen C, Sunjevaric I, Rothstein R. DNA strand annealing is promoted by the yeast Rad52 protein. *Proc Natl Acad Sci U S A.* Oct 1.1996 93:10729. [PubMed: 8855248]

53. 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.* Sep.2006 38:1043. [PubMed: 16921376]
54. Galanty Y, et al. Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. *Nature.* Dec 17.2009 462:935. [PubMed: 20016603]
55. Galanty Y, Belotserkovskaya R, Coates J, Jackson SP. RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev.* Jun 1.2012 26:1179. [PubMed: 22661229]
56. Liauw SL, Connell PP, Weichselbaum RR. New paradigms and future challenges in radiation oncology: an update of biological targets and technology. *Sci Transl Med.* Feb 20.2013 5:173sr2. [PubMed: 23427246]

Figure 1. Pathways and genes involved in repair of double-strand DNA breaks (DSBs) and the tolerance of replication stress

Shown is a simplified overview of the mechanistic steps and genes involved in DNA repair, with an emphasis on those that facilitate homologous recombination and non-homologous end joining. All of the displayed genes were considered candidates for the Recombination Proficiency Score (RPS) system, except those within the blue box. The four genes whose expression levels were ultimately chosen to comprise the RPS are displayed in red.

Figure 2. Cell lines with low RPS overexpress a wide array of HR-related genes

Mean mRNA levels are shown for the CCLE cell lines with low RPS scores. These mRNA levels were mined from the CCLE database, and displayed values represent log₂ transformed mRNA measurements of each gene normalized to the median mRNA among the starting 634 carcinoma cell lines. Therefore an expression level of zero indicates a median expression level, and any positive value indicates overexpression. For example, a value of +0.25 indicates a 19% increase in expression above the median. Error bars denote standard error.

Figure 3. RPS predicts sensitivity to different classes of treatment and predicts HR deficiency in cell lines

A) CCLE carcinoma cell lines were binned into quartiles, based on RPS. Sensitivity data were mined from the CCLE database and plotted for different oncologic therapies, and differences between the highest and lowest quartiles were determined by Student's T test. **B)** HR repair efficiency correlates with RPS. Six representative cell lines were co-transfected with an HR reporter-containing plasmid (pDR-GFP) plus an I-Sce I expressing plasmid (pCβASce) or an empty vector control plasmid (pCAG), and were subjected to FACS analysis 48 hours later. Reported HR efficiency represents the percent GFP+ cells with pDR-GFP + pCβASce, normalized to background (pDR-GFP + pCAG).

A) TP53 wild-type cell lines **B) TP53 mutant cell lines**

Figure 4. CCLE carcinoma cell lines with low RPS have elevated genomic instability
 SNP array-based DNA copy number variations (CNVs) were mined from the CCLE database. DNA deletions (left) and amplifications (right) were binned by size, wherein bins represent 10-fold increments in mutation size. High and low RPS groups were defined as the top and bottom quartiles, respectively. Size-based distributions of CNVs are shown for **A)** TP53 WT cells and **B)** TP53 mutant cells. Error bars denote standard error. Asterisks denote significant differences, based on Student's T test.

a-b

Figure 5a

c-d

Figure 5b**Figure 5. Low RPS is associated with genomic instability in human tumors**

SNP array-based DNA copy number variations (CNVs) were mined from the Cancer Genome Atlas. DNA deletions (left) and amplifications (right) were binned by size, wherein bins represent 10-fold increments in size. High and low RPS groups were defined as the top and bottom quartiles, respectively. Size-based distributions of CNVs are shown for **A)** TP53 WT NSCLC tumors, **B)** TP53 mutant NSCLC tumors, **C)** TP53 WT breast tumors, **D)** TP53 mutant breast tumors. Error bars denote standard error. Asterisks denote significant differences, based on Student's T test.

Figure 6. RPS is prognostic and predictive of treatment sensitivity in clinical tumors

A) Kaplan Meier survival curves are shown for NSCLC patients treated on the JBR.10 trial with either surgery alone (S) or surgery followed by chemotherapy (S+C). Low and high RPS groups were defined as the bottom 25th percentile and the remaining upper 75th percentile, respectively. **B)** Four clinical datasets of non-small cell lung cancer were analyzed for prognostic impact of RPS on survival, using multivariate analyses that controlled for overall stage. Points in the Forest plot represent treatment-specific hazard ratios of RPS (as a continuous variable). Boxes denote hazard ratio and diamonds denote modeled hazard ratio values that summarize the combined impact of all four datasets. Error bars denote 95% confidence intervals. Black= surgery alone, green= surgery + chemotherapy.

Table 1
Low RPS is associated with adverse clinical features in human tumors

Prognostic Factor	RPS Quartile				p-value
	0-25th	25-50th	50-75th	75-100th	
Non-small cell lung cancer					
T3/4 tumor	16%	11%	14%	8%	0.75
Lymph node invasion	51%	33%	33%	14%	0.0085
TP53 mutation	89%	75%	86%	44%	<0.0001
Breast cancer					
T3/4 tumor	11%	16%	10%	16%	0.66
Lymph node invasion	48%	59%	58%	46%	0.29
TP53 mutation	60%	40%	39%	19%	<0.0001
Estrogen receptor loss	46%	23%	17%	6%	<0.0001
HER2 amplification	14%	19%	19%	0%	0.0072

p-values denote differences in frequencies among groups based on a likelihood ratio test