Identifying DNA Repair Enzymes as Potential Therapeutic Targets in Gastrointestinal Tumors: A Drug Repurposing Study

Kiran Tikoo

Received November 11, 2024
Accepted March 02, 2025
Electronic access March 15, 2025

Using publicly available gene expression datasets, several DNA damage response (DDR) genes were identified in a subset of gastrointestinal cancers which were significantly overexpressed, and then targeted for inhibition by repurposed drugs in silico. Initially, it was found that 11 out of 217 DDR genes (MSH2, RAD51, UBE2T, FEN1, EXO1, PRKDC, H2AX, CHEK1, FANCI, SHFM1, and PCNA) were commonly overexpressed across 5 GI Cancer types via GEPIA2 differential expression analysis, including colon adenocarcinoma (COAD), esophageal adenocarcinoma (ESCA), stomach adenocarcinoma (STAD), pancreatic adenocarcinoma (PAAD), and liver hepatocellular adenocarcinoma (LIHC). It was found that in LIHC, expression of 4 of these 11 DDR genes (EXO1, RAD51, FANCI, and CHEK1) A) were highly correlated with each other (defined as pairs of genes with a Pearson correlation coefficient of 0.6 or greater and a p value of less than or equal to 0.05), B) were significantly detrimental for patient survival (defined as a hazard ratio greater than 1.5 and a p value cutoff of 0.05), and C) displayed minimal protein and RNA expression in normal tissues, all of which suggested that they may be good targets for new therapies. Two of the DDR genes (RAD51 and FANCI) fit these same three criteria in PAAD. Using a molecular-screening strategy, it was discovered that Saquinavir, an anti-HIV drug, bound to the protein RAD51 in silico (Vina score -9) with a higher binding strength than existing, previously published RAD51 inhibitory molecular probes. It was also discovered that Ridinilazole, an antibiotic, bound to the protein EXO1 in silico (Vina score -9.1) with a comparable binding affinity to published EXO1 inhibitory probes. While further biological studies are necessary, this bioinformatics study suggests that Saquinavir and Ridinilazole could be repurposed to inhibit RAD51 and EXO1, respectively, and potentially treat specific gastrointestinal cancers.

Keywords: DNA Damage Repair, Gastrointestinal Tumors, Bioinformatics Analysis, Drug repurposing

Introduction

Gastrointestinal (GI) cancers are among the most malignant tumors. With an incidence of 4.8 million and mortality of 3.4 million patients, these cancers constituted 26% of new cancer cases and 35% of cancer-related deaths in 2018[1]. By 2040, the incidence and mortality of GI Cancers are predicted to increase to 7.5 and 5.6 million respectively.

A majority of gastrointestinal cancers are caused by modifiable risk factors such as smoking and alcohol consumption[1]. With colorectal cancers as an exception, the mortality for GI cancers is about 1.4 to 1.8 times greater than GI cancer incidence. Nearly all gastrointestinal tumors are difficult to diagnose early. Late detection then leads to a generally poor prognosis. Therefore, better screening and treatment options are needed for GI tumors in order to improve outcomes.

Since genomic instability is a hallmark of cancer, understanding the regulation of DNA repair pathways is necessary to treat cancer[2]. While it has long been understood that defects in DNA repair can predispose to cancer, DDR gene expression in cancer can also become upregulated to help ensure tumor survival against radiotherapy and chemotherapy, a process known as non-oncogene addiction[3,4]. Such genes may represent critical bottlenecks for tumor survival, and represent promising targets for new treatments, such as inhibitors of DDR genes, or inhibitors of their synthetic lethal pathways that are related to their functioning[3].

Currently, there is a paucity of FDA approved drugs to target the DDR pathways. Despite the large numbers of DDR genes (over 200), the only clinically-approved treatments are PARP inhibitors. Given this disparity, an investigation of 219 DDR genes across multiple pathways was undertaken to evaluate their potential drug targetability. Following a multi-faceted approach using patient-derived and cell-derived datasets in combination with bioinformatics analysis, 4 potential DDR gene targets were identified for COAD, ESCA, LIHC, PAAD, STAD, and their potential inhibitors were identified.

Methods

The 219 DDR genes analyzed in this study were identified from the Richard D. Wood lab website at MD Anderson[5].

Analysis of gene expression was conducted utilizing GEPIA2 (gepia2.cancer-pku.cn). GEPIA2 develops new computational tools derived from the recomputed TCGA and GTEx data of UCSC[6]. COAD, ESCA, LIHC, PAAD, and STAD were investigated using GEPIA2's differential expression analysis tool. GEPIA2 quantifies gene expression by the measure of Log_2FC only for genes significantly upregulated or downregulated. The default parameters were used for the GEPIA2 differential expression analysis (Differential Method = ANOVA, Gene/Isoform = Gene, $|Log_2FC|$ Cutoff = 1, q-value Cutoff = 0.01, Chromosomal Distribution = Both). The search menu in the list feature was used to identify the Log_2FC for each gene when available for each cancer type. When available from GEPIA2, the Log_2FC values were recorded for these DDR genes in the following cancers: COAD, ESCA, LIHC, PAAD, and STAD.

The survival analysis feature was utilized from GEPIA2 to analyze survival across all cancer types. A Kaplan-Meier chart was generated for each individual DDR gene for each cancer type (COAD, ESCA, LIHC, PAAD, STAD). The default settings were used (Method = Overall Survival, Hazards Ratio = Yes, 95% Confidence Interval = Yes, Axis Units = Months). The dataset was the specific cancer type in question. Two DDR genes (DNTT and SPO11) were excluded from survival analysis as there was no available data in GEPIA2. Hazard ratios were recorded, with significant results identified ($p \leq 0.05$).

Another GEPIA2 survival analysis was conducted for PAAD and LIHC cancer subtypes with the same settings as the initial analysis (Method = Overall Survival, Hazards Ratio = Yes, 95% Confidence Interval = Yes, Axis Units = Months). All resulting hazard ratios were recorded, and significant results (where the p value was less than or equal to 0.05) were identified.

Correlated expression of DDR genes in LIHC and PAAD were determined using the data from GEPIA2 boxplot expression. The GEPIA2 correlation analysis feature was used with the default settings (Correlation Coefficient = Pearson). The expression datasets used were LIHC Tumor and PAAD Tumor. Every combination of two genes with a HR greater or equal to 1.5 were tested (5 genes for PAAD and 9 for LIHC). A significant Pearson correlation coefficient value was considered to have a p value of less than or equal to 0.05. The significant Pearson correlation coefficient values were recorded for every combination of two genes from each tumor.

A set of 7 strongly correlated genes for LIHC and 4 genes for PAAD (Pearson Correlation Coefficient ≥ 0.6 between any two genes in the signature) were used to develop a gene signature, and the hazard ratios for these two signatures were evaluated through the survival analysis feature from GEPIA2.

These two gene signatures underwent a STRING (string-db.org) analysis to determine whether the genes produce proteins that are physically or functionally connected[7]. The Multiple Proteins feature was used in the STRING database utilizing the default settings (Network Type: full STRING network, Required score: medium confidence (0.400), FDR stringency: medium (5 percent)). The networks and corresponding PPI enrichment p-value scores were recorded.

These 11 DDR genes (7 for LIHC and 4 for PAAD) underwent further analysis using shinyDepMap (labsyspharm.shinyapps.io/depmap)[8]. Under gene essentiality, each gene from the list was tested (PRKDC, PARPBP, RECQL4 and SLX1B were excluded from analysis due to lack of data). The selectivity and efficacy scores of each gene were recorded.

RNA and protein expression in normal tissue was evaluated for these 11 DDR genes using The Human Protein Atlas(proteinatlas.org)[9]. The Protein Expression Overview and the Consensus Dataset from the RNA Expression Overview were utilized. For the protein tissue expression analysis, the number of tissues that expressed undetected, low, medium, and high protein levels were counted for each DDR protein. For the RNA tissue expression analysis, the number of tissues that expressed 0-5 normalized transcripts per million (nTPM), 5.1-10 nTPm, 10.1-20 nTPM, 20.1-50 nTPM, 50.1-100 nTPM, and greater than 100 nTPM were counted.

The synthetic lethal matches for each DDR gene were identified using the search feature of the newest version of SynLethDB (SynLethDB 3.0) (synlethdb.com)[10]. All synthetically lethal pairs were recorded with the search criteria of a Statistic Score less than or equal to 0.500. These gene pairs were then input into GEPIA2 as signatures in a survival analysis in both PAAD and LIHC using detail settings (Method = Overall Survival, Hazards Ratio = Yes, 95% Confidence Interval = Yes, Axis Units = Months).

Potential small molecule inhibitors of each of these four proteins were identified using literature searches. FANCI had no direct inhibitors that could be found, thus it was excluded from further analysis. The SwissSimilarity (swisssimilarity.ch) tool was used to identify chemically similar inhibitors to these published inhibitors[11]. The PubChem (pubchem.ncbi.nlm.nih.gov) database was used to find the canonical SMILES for each published inhibitor as the input for SwissSimilarity[12]. Compounds selected for further analysis were selected from the Drug Bank dataset using the criteria of a SwissSimilarity score of greater than 0.7.

To determine the binding affinity of the potential inhibitors identified through SwissSimilarity, CB-Dock2 (cadd.labshare.cn/cb-dock2) was utilized[13]. The protein structures for RAD51, CHEK1, FANCI, and EXO1 were obtained from RCSB (rcsb.org) in the PDB file format[14]. The Nucleic Acid Knowledgebase identifiers (NAKB) of the RAD51 protein structure from RCSB was 8bpc. The CHEK1 protein structure was downloaded from RCSB (the NAKB was 4fsm). The EXO1 DNA complex was obtained from RCSB (the NAKB was 3qe9). The 3D SDF files of all potential inhibitors identified from SwissSimilarity and the published inhibitors were all obtained from Pubchem: However, the SDF for Saquinavir was

unavailable at Pubchem, so the shape of the drug was obtained from RCSB (the NAKB was 6xcz). The ligand inputs for CB-Dock2 were each of the inhibitor SDF files. The protein inputs for CB-Dock2 were the PDB protein files for each of the four DDR proteins. The outputs for CB-Dock2 included a binding affinity score (Vina score) at each of five different inhibitor-protein binding sites (labeled C1, C2, C3, C4, C5). Out of the five possible inhibitor-binding sites, the assumed biological site was identified as the site to which the chemical probe for the protein had the highest binding affinity.

The three dimensional structure of each of these five inhibitor-protein binding sites were visualized using PyMol (pymol.org)[15]. The drug-protein molecular interactions for the assumed binding sites for each protein and drug combination were then analyzed using Arpeggio (biosig.lab.uq.edu.au/arpeggioweb) when possible[16]. All heteroatom groups and chain-everything interactions were selected as entities to calculate contacts. The resulting PyMol session files containing protein-drug structure were then analyzed to contrast the structural differences between the published literature and chemically similar drugs.

Results

Initially, the presence of non-oncogene addiction of DDR genes in GI cancers was investigated by evaluating their expression levels through GEPIA2. From the 219 DDR genes evaluated, the expression of 136 of these genes was significantly upregulated ($Log_2FC > 1$ and $p \leq 0.01$) in at least one GI cancer type (COAD, ESCA, LIHC, PAAD, or STAD). It was found that 6 genes were significantly upregulated in all tumors except ESCA, 11 genes were upregulated in all tumors except LIHC, and one gene was upregulated in all genes except COAD (See Figure 1, left panel). The Venn diagram in Figure 1, left panel, further details the sets of genes commonly upregulated among these five GI tumor types.

Significantly, 11 genes were commonly overexpressed in each of the five GI cancer types: MSH2, RAD51, UBE2T, FEN1, EXO1, PRKDC, H2AX, CHEK1, FANCI, SHFM1, PCNA (see Figure 1, right panel). Log_2FC values ranged from 1.062 to 3.423 among these genes. Involved pathways included mismatch excision repair, homologous recombination, fanconi anemia, non-homologous end-joining, editing and processing nucleases, and chromatin structure and modification.

Next, it was hypothesized that these overexpressed genes were important to promote cancer survival, so the association between gene expression and patient survival was addressed. Using GEPIA2's survival analysis tool, all hazard ratios were recorded for these 11 DDR genes, among each of the 5 cancer types. Only LIHC and PAAD contained DDR genes with a significant hazard ratio greater than 1 (with p value ≤ 0.05), signifying worse patient survival. In LIHC, 9 genes exhibited significantly elevated hazard ratios from 1.5 to 2.1: MSH2, RAD51, UBE2T, FEN1, EXO1, PRKDC, H2AX, CHEK1, and FANCI (see Figure 2A). In PAAD, 5 genes exhibited significantly elevated hazard ratios from 1.7 to 1.9: MSH2, RAD51, FANCI, UBE2T, and FEN1.

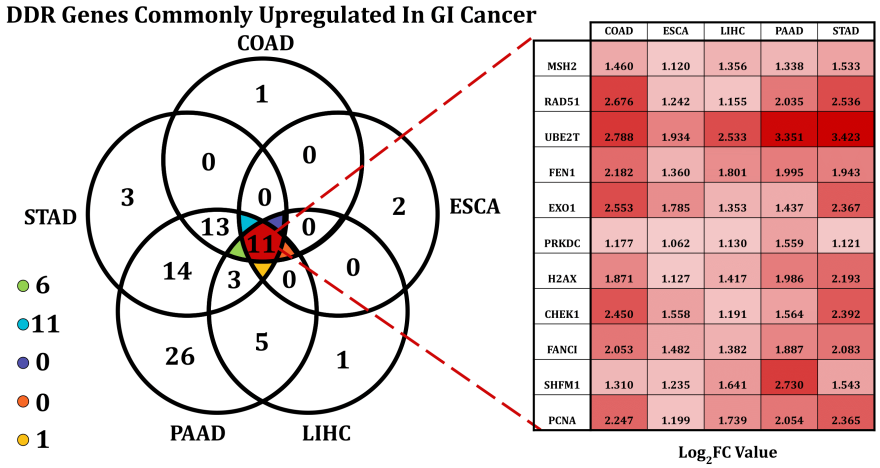
Hazard ratios represent the relative likelihood of a disease being resolved in the control group relative to the test group: for example, a hazard ratio (HR) of 2 indicates that a patient in the control group is twice as likely to have the disease resolved than a patient in the testing group. In the GEPIA2 survival analyses, the control group contains below median expression of a gene, while the test group contains above median expression[17]. A DDR gene with $HR > 1$ in a particular cancer indicates that this gene is detrimental to the disease resolution (of cancer), and suggests that the gene plays a biologically significant role in disease progression.

Since DDR genes act along several pathways which may be synergistic in DNA damage repair, an attempt was made to identify correlated DDR gene expression and its impact on cancer survival. Therefore these 9 genes for LIHC and 5 genes for PAAD underwent further correlation analysis. The Pearson correlation coefficients were evaluated for each combination of two genes in the above LIHC and PAAD gene sets using GEPIA2's correlation analysis tool. In this analysis, the Pearson correlation coefficient values ranged from 0.45 to 0.84 in the LIHC gene sets and from 0.41 to 0.81 in the PAAD gene sets (see Figure 2B, C).

Further investigation was conducted to identify whether a subset of these genes had synergistic effects on patient survival. A signature for each LIHC and PAAD was created that contained only commonly overexpressed genes with high hazard ratios that were strongly correlated with one another. A Pearson correlation coefficient of ≥ 0.6 was defined as a strong correlation between two genes. From this criteria, two gene signatures were created. LIHC's gene set included MSH2, RAD51, FANCI, UBE2T, FEN1, EXO1, and CHEK1 (see Figure 2B). PAAD's gene set included MSH2, RAD51, FANCI, UBE2T, and FEN1 (see Figure 2C).

These gene signatures were also validated using STRING. The interaction enrichment in each signature was quantified using the PPI enrichment p-value. The lower p-value, the more likely it is that there is a biological connection between the proteins in the set. The PPI enrichment p-value was 1.58e-14 for LIHC and 1.95e-5 for PAAD. These p-values were much lower than required ($p \leq 0.05$) to indicate a biologically significant connection between the proteins in the set (see Figure 2B, C).

The hazard ratio for the LIHC signature (MSH2, RAD51, FANCI, UBE2T, FEN1, EXO1, and CHEK1) was 1.8, while the HR for the individual genes ranged between 1.5 and 2.1. The hazard ratio for the PAAD signature (MSH2, RAD51, FANCI, UBE2T, and FEN1) was 1.5, while the HR for the individual genes ranged from 1.7 to 1.9. For both LIHC and PAAD, the



	COAD	ESCA	LIHC	PAAD	STAD
MSH2	1.460	1.120	1.356	1.338	1.533
RAD51	2.676	1.242	1.155	2.035	2.536
UBE2T	2.788	1.934	2.533	3.351	3.423
FEN1	2.182	1.360	1.801	1.995	1.943
EXO1	2.553	1.785	1.353	1.437	2.367
PRKDC	1.177	1.062	1.130	1.559	1.121
H2AX	1.871	1.127	1.417	1.986	2.193
CHEK1	2.450	1.558	1.191	1.564	2.392
FANCI	2.053	1.482	1.382	1.887	2.083
SHFM1	1.310	1.235	1.641	2.730	1.543
PCNA	2.247	1.199	1.739	2.054	2.365

Log_2FC Value

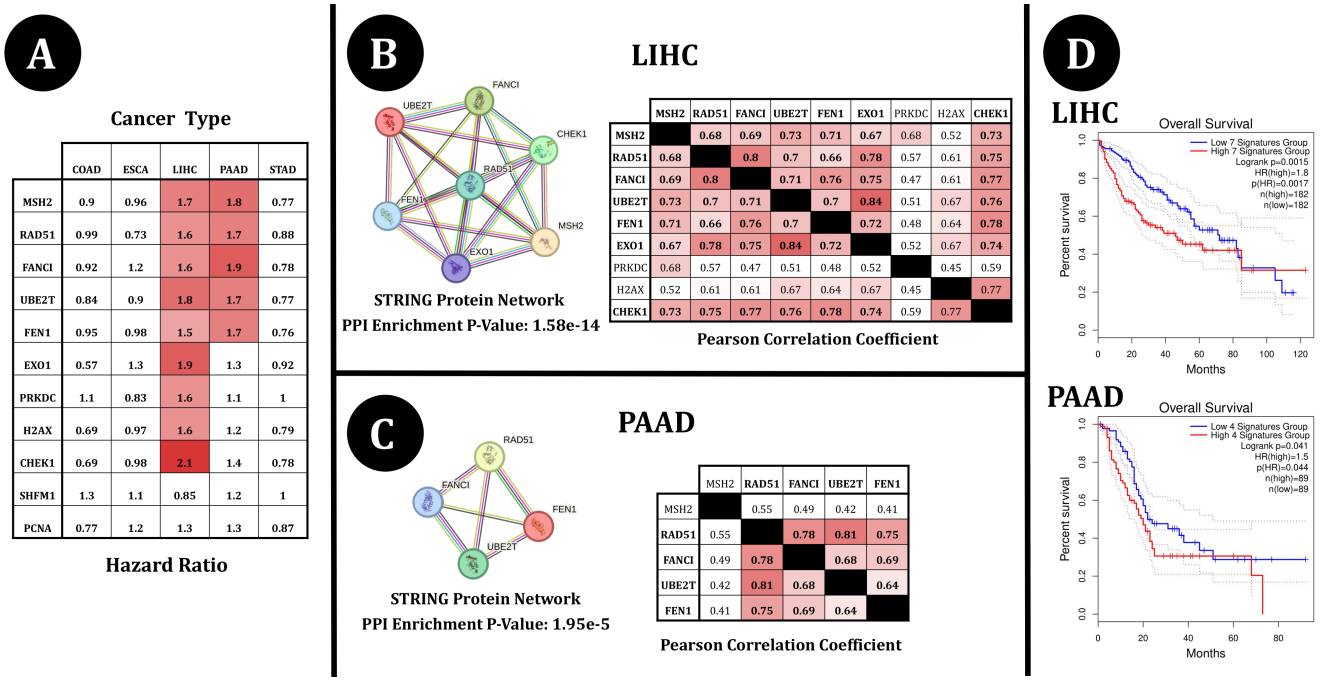
Fig. 1 Left panel: Venn Diagram depicting the number of commonly overexpressed genes, in each GI cancer type. 11 DDR genes were upregulated in each of 5 GI cancer types (in red at the center). See text for further details. Right Panel: Corresponding Log_2FC values ($p \leq 0.05$) for the 11 upregulated genes in each cancer type. COAD: Colorectal Adenocarcinoma, ESCA: Esophageal Adenocarcinoma, LIHC: Liver Hepatocellular Adenocarcinoma, PAAD: Pancreatic Adenocarcinoma, STAD: Stomach Adenocarcinoma. Darker red represents higher Log_2FC values, while light red represents lower Log_2FC values).

hazard ratios of the signatures were not significantly higher than the hazard ratios of the individual genes. This meant that the combined effect of these signatures did not have an additive effect on survival.

Our initial set of 11 commonly overexpressed DDR genes underwent further investigation with the Human Protein Atlas to identify the best potential gene target for future therapies. While it was already demonstrated that all these genes were commonly overexpressed in 5 GI cancers, these genes should also be expressed at low levels in normal tissue to avoid any off-target side-effects, if these genes are to be considered good therapeutic targets. The RNA tissue distribution of each of these 11 genes was determined in each of 50 normal tissues from the Human Protein Atlas (see Figure 3A). RNA expression levels were quantified in nTPM (normalized transcripts per million) and were binned into the following categories: 0-5 nTPM, 5.1-10 nTPM, 10.1-20 nTPM, 20.1-50 nTPM, 50.1-100 nTPM and >100 nTPM. A large number of normal tissues had low levels of RNA expression of EXO1, RAD51, CHEK1, and FANCI (45, 41, 37, and 31 tissues in the 0-5 nTPM category respectively, see Figure 3A, asterisks). This suggested that these genes could be good therapeutic targets, as their inhibitors may avoid significant off-target effects. The remaining genes, including MSH2, UBE2T, FEN1, PRKDC, H2AX, SHFM1, and PNCA, displayed significantly higher RNA expression in most normal tissues, suggesting that they would not be optimal therapeutic targets, as their potential inhibitors may display significant off-target effects.

The Human Protein Atlas also contained protein expression data for 6 of these 11 genes. Protein expression levels were obtained of MSH2, RAD51, FEN1, H2AX, SHFM1, and PCNA in 44-46 normal tissues. Tissue expression levels are designated as either undetected, low, medium, or high, by the Human Protein Atlas, by identifying the frequency and intensity of antibodies binding to tissue samples [10]. Of these 6 genes, RAD51 was the only one with predominantly low protein expression in normal tissues: 16 tissues had no detectable RAD51 protein, and 25 had low RAD51 protein expression (see Figure 3B). Of the remaining tissues, only 3 tissues (ovary, skin, and tonsil) had medium expression, and only 2 tissues (testis and thymus) had high expression of RAD51. MSH2, FEN1, H2AX, SHFM1, and PCNA had predominantly medium or high protein expression in most tissues, confirming that they would not be optimal therapeutic targets because their potential inhibitors may display significant off-target effects.

To corroborate these results, ShinyDepMap was utilized in order to evaluate the efficacy and selectivity of these genes within cancer cell lines. A good drug target should have high selectivity and substantial efficacy (the more substantial, the more negative efficacy score) within cancer cell lines. The efficacy measure is defined as how much the loss of a gene impacts cell growth, and the selectivity measure is defined as how much efficacy varies across multiple cell lines. Genes that represent ideal therapeutic targets will have high efficacy and selectivity. The most promising DDR targets from the Human Protein Atlas expression analysis, EXO1, RAD51, CHEK1, and FANCI, had efficacy scores of -0.478, -1.51, -1.666, and -0.582, respectively, and selectivity scores of 0.25, 0.163, -0.002, and 0.126, respectively (see Figure 3C). The efficacy scores of these four genes suggest that they promote cell growth in cancer



	COAD	ESCA	LIHC	PAAD	STAD
MSH2	0.9	0.96	1.7	1.8	0.77
RAD51	0.99	0.73	1.6	1.7	0.88
FANCI	0.92	1.2	1.6	1.9	0.78
UBE2T	0.84	0.9	1.8	1.7	0.77
FEN1	0.95	0.98	1.5	1.7	0.76
EXO1	0.57	1.3	1.9	1.3	0.92
PRKDC	1.1	0.83	1.6	1.1	1
H2AX	0.69	0.97	1.6	1.2	0.79
CHEK1	0.69	0.98	2.1	1.4	0.78
SHFM1	1.3	1.1	0.85	1.2	1
PCNA	0.77	1.2	1.3	1.3	0.87

	MSH2	RAD51	FANCI	UBE2T	FEN1	EXO1	PRKDC	H2AX	CHEK1
MSH2		0.68	0.69	0.73	0.71	0.67	0.68	0.52	0.73
RAD51	0.68		0.8	0.7	0.66	0.78	0.57	0.61	0.75
FANCI	0.69	0.8		0.71	0.76	0.75	0.47	0.61	0.77
UBE2T	0.73	0.7	0.71		0.7	0.84	0.51	0.67	0.76
FEN1	0.71	0.66	0.76	0.7		0.72	0.48	0.64	0.78
EXO1	0.67	0.78	0.75	0.84	0.72		0.52	0.67	0.74
PRKDC	0.68	0.57	0.47	0.51	0.48	0.52		0.45	0.59
H2AX	0.52	0.61	0.61	0.67	0.64	0.67	0.45		0.77
CHEK1	0.73	0.75	0.77	0.76	0.78	0.74	0.59	0.77	

	MSH2	RAD51	FANCI	UBE2T	FEN1
MSH2		0.55	0.49	0.42	0.41
RAD51	0.55		0.78	0.81	0.75
FANCI	0.49	0.78		0.68	0.69
UBE2T	0.42	0.81	0.68		0.64
FEN1	0.41	0.75	0.69	0.64	

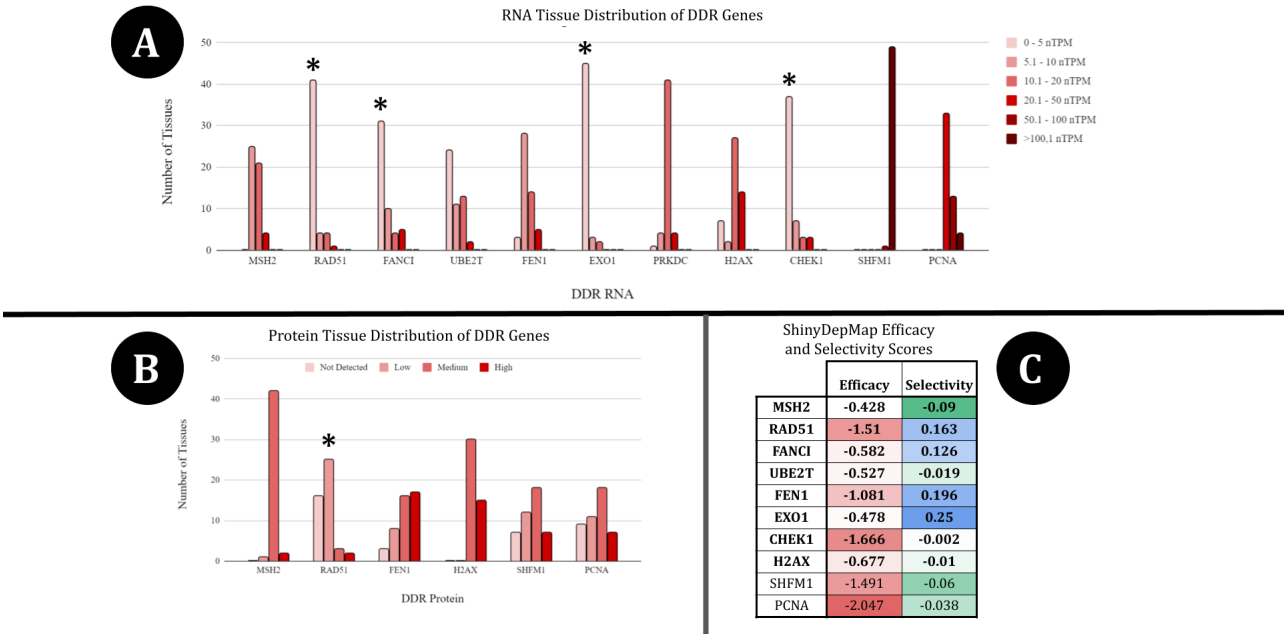
Fig. 2 (A) Hazard ratios among the 11 commonly overexpressed genes in each of 5 GI cancer types (significant HR >1 and $p \leq 0.05$ are shown in red). (B) Pearson correlation coefficient values between any two genes with a significant HR >1 in LIHC ($p \leq 0.05$). Darker red indicates a higher Pearson Correlation Coefficient. STRING Protein Network Interactions also shown. The line color between proteins indicates the evidence used to link two proteins [8]. Red lines indicate the presence of fusion evidence, green lines indicate neighborhood evidence, blue lines indicate cooccurence evidence, purple lines indicate experimental evidence, yellow lines indicate text mining evidence, light blue lines indicate database evidence, and black lines indicate coexpression evidence. PPI Enrichment P-Value is also shown: a low ($p \leq 0.05$) PPI Enrichment P-Value indicates that the proteins in the network are biologically connected. (C) Pearson correlation coefficient values between any two genes with a significant HR >1 in PAAD ($p \leq 0.05$). Darker red indicates a higher Pearson Correlation Coefficient. STRING Protein Network Interactions and PPI Enrichment P-Value are also shown, as above (D) Kaplan-Meier charts for the identified gene signatures for LIHC and PAAD (see text for details).

cell lines, and could represent good therapeutic targets to treat cancer.

Given (1) their elevated expression in multiple GI cancers, (2) their significant negative impact on survival, (3) their significant correlated expression, (4) their lack of expression in most normal tissues, and (5) their positive impact on growth of cancer cell lines, potential inhibitors of RAD51, CHEK1, FANCI, and EXO1 remained a promising pathway for further inquiry as a treatment for GI cancers. Subsequently, a strategy was devised to identify repurposed FDA approved drugs that could inhibit these proteins and potentially treat GI cancers.

Existing published inhibitors were first identified for RAD51, CHEK1, FANCI, and EXO1 through literature searches. The SwissSimilarity tool was used to identify small molecule inhibitors chemically similar to these published inhibitors, using the DrugBank option to search for similar FDA-approved drugs. The Canonical SMILES format for each published inhibitor was used in SwissSimilarity as the search criteria within the DrugBank dataset. For RAD51, 3 published inhibitors (RI-1, CAM833, and B02) were identified from literature searches, with 6 similar molecules found through SwissSimilarity (with a SwissSimilarity score of >0.7). Only one of these similar molecules, Saquinavir, had binding properties (see below) to suggest that it would be a potential inhibitor of RAD51. For CHEK1, 3 published inhibitors (CHIR-124, 7-hydroxystaurosporine, and SB-218078) were investigated with 5 similar molecules found. One of these similar molecules, Lestaurtinib, displayed binding properties to suggest that it would be a potential inhibitor of CHEK1. For EXO1, 1 published inhibitor was found and 1 similar molecule was found, Ridinilazole [19]. No published inhibitors were found for FANCI, so it was excluded from further analysis.

To further investigate these inhibitors CB-Dock2 was used to find the binding affinities of each inhibitor with their target DDR protein. The binding affinity of each inhibitor in each of five potential binding sites was measured by using the Vina score provided by CB-Dock2. Out of these five potential binding sites displayed by CB-Dock2, the primary binding site was defined as the site with the strongest Vina score among all the published inhibitors.



	Efficacy	Selectivity
MSH2	-0.428	-0.09
RAD51	-1.51	0.163
FANCI	-0.582	0.126
UBE2T	-0.527	-0.019
FEN1	-1.081	0.196
EXO1	-0.478	0.25
CHEK1	-1.666	-0.002
H2AX	-0.677	-0.01
SHFM1	-1.491	-0.06
PCNA	-2.047	-0.038

Fig. 3 (A) RNA tissue distribution for 11 DDR genes in normal tissue. Low RNA expression of RAD51, FANCI, EXO1, and CHEK1 is present in most normal tissues (asterisks). The Human Protein Atlas data has error in 0.1% to 0.01% of RNA sequencing data due to the Illumina RNA sequencing method used [10,18]. Darker red signifies higher RNA expression. The y-axis represents the numbers of tissues with the indicated expression levels specified in nTPM in the legend. (B) Protein expression levels for 6 DDR genes in normal tissue. Low protein expression was found for RAD51 in most tissues (asterisk). Darker red signifies higher protein expression. The y-axis represents the numbers of tissues with the indicated expression levels. (C) shinyDepMap Efficacy and Selectivity data among 10 of the 11 overexpressed genes: darker colors represent extreme values.

In the analysis of CHEK1, SwissSimilarity revealed that the inhibitor Lestaurtinib was chemically similar to the published inhibitor 7-hydroxy-staurosporine [20] (SwissSimilarity score of 0.980, see Figure 4). The CHEK1 binding affinity to Lestaurtinib as determined by CB-Dock2 (Vina score of -10.2, see Figure 4B) was also comparable to its binding affinity to 7-hydroxy-staurosporine (Vina score of -9.8, see Figure 4A).

To determine the specific mechanisms of inhibitor-protein binding, the specific molecular interactions between CHEK1 and 7-hydroxy-staurosporine or Lestaurnib were analyzed using Arpeggio (see Figure 4). The strongest chemical bonds identified from this analysis were CARBON PI and polar hydrogen bonds. There were 6 CARBON PI interactions present between Lestaurtinib and CHEK1-Valine 23, while there were only 5 CARBON PI present between 7-hydroxy-staurosporine and Valine 23. This additional CARBON PI interaction between Lestaurtinib and Valine 23 that was not present in 7-hydroxy-staurosporine may be one determinant of the stronger in silico binding between Lestaurtinib and CHEK1 (see Figure 4B). In addition, there was one polar hydrogen bond between 7-hydroxy-staurosporine and CHEK1-Glutamate 85, and this bond was not present with Lestaurnib. Furthermore, there were 2 CARBON PI interactions for each 7-hydroxy-staurosporine and Lestaurtinib with CHEK1-Leucine 137. This similarity of

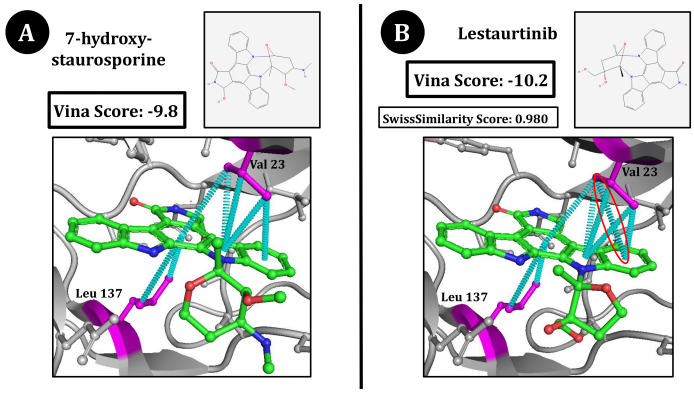


Fig. 4 (A) Chemical structure, Vina score, and Arpeggio binding data for the published inhibitor 7-hydroxy-staurosporine bound to CHEK1 at binding site C2. Carbon PI bonds are shown in cyan and contact residues are shown in magenta. (B) Chemical structure, Vina score, SwissSimilarity score, and Arpeggio binding data for the chemically similar inhibitor Lestaurtinib bound to CHEK1 at binding site C2. Carbon PI bonds are shown in cyan and contact residues are shown in magenta. The additional Carbon PI bond between Lestaurtinib and CHEK1-Valine-23 is circled in red.

the molecular interactions between each inhibitor (Lestaurtinib and 7-hydroxy-staurosporine) with CHEK1 suggests that Lestau-

rtinib may inhibit CHEK1, similar to 7-hydroxy-staurosporine.

In the analysis of EXO1 inhibitors, the chemically similar inhibitor Ridinilazole (SwissSimilarity score of 0.948) had relatively comparable binding affinity as the published inhibitor CID_824226. Out of the 5 potential binding sites identified by Cb-Dock2, site C1 was defined as the primary binding site because it had the strongest binding affinity with the published inhibitor CID_824226. At this site, CID_824226 and Ridinilazole had comparable binding affinities (Vina scores of -9.4 and -9.1 respectively, see Figure 5). These similar binding affinities at the same site suggest that Ridinilazole could potentially inhibit EXO1.

Analysis of the RAD51 inhibitors revealed one chemically similar inhibitor, Saquinavir (SwissSimilarity score of 0.717), that had a stronger binding affinity to RAD51 compared to the published inhibitors. From the published inhibitors, the binding pocket with the strongest binding affinity varied. Saquinavir had a binding affinity of -7.9 and -8.4 at the sites C1 and C3, respectively, which were also the sites with the most substantial binding affinity in the published inhibitors RI-1 (Vina score -6.5) and B02 (Vina score -7.4) respectively (see Figure 6). Saquinavir also had a somewhat comparable binding affinity to the published inhibitor CAM833 in site C5 (Vina score of -7.7 compared to -8.4, respectively). These data suggest that Saquinavir may serve as a potential inhibitor of RAD51.

Synthetic lethal analysis was then undertaken to identify additional drug targets. Using SynLethDB, RAD51, CHEK1, FANCI, and EXO1 underwent synthetic lethal analysis in LIHC, with 70 potential synthetic lethal gene pairs found. RAD51 and FANCI underwent synthetic lethality analysis in PAAD, with 16 potential synthetic lethal gene pairs found. The hazard ratio of these DDR genes individually was compared to the hazard ratio for the synthetic lethal pair. In LIHC, the hazard ratio of the 53 synthetic lethal CHEK1 gene pairs ranged from 1.1 to 2.1, while the individual HR for CHEK1 was 2.1. In addition, in LIHC, the hazard ratio of the 15 synthetic lethal RAD51 gene pairs ranged from 1.4 to 1.9, while the HR for RAD51 was 1.6. For PAAD, the hazard ratio of the 15 synthetic lethal RAD51 gene pairs ranged from 1.3 to 1.9 while the HR for RAD51 was 1.7. All synthetic lethal pairs only had a marginal (0.1-0.3 increase in HR) change in survival. Considering that synthetically lethal genes should lead to complete deactivation of a pathway, large changes in hazard ratio were expected. Therefore, it was concluded that these synthetic lethal gene pairs were unlikely to be biologically significant and therefore were unlikely to be good drug targets.

Since tumor subtypes can have significant impacts on therapy, a survival analysis of the most promising DDR targets, RAD51 and EXO1, with subtypes of LIHC and PAAD was undertaken through GEPIA2 survival analysis. This analysis revealed high hazard ratios for EXO1 (HR = 2.2) in basal pancreatic ductal adenocarcinoma and in iCluster 1 and 2 for LIHC (HR = 2.9 and HR = 2.2 respectively, see Figure 7). The iCluster framework is used to determine cancer subtypes using genomic data and iCluster 1, 2, and 3 were used in a subtype analysis for LIHC in GEPIA2[21]. A high hazard ratio was also observed for RAD51 in LIHC iCluster 1 (HR = 2.2). This suggests that inhibition of RAD51 with Saquinavir or EXO1 with Ridinilazole may be more effective in these specific subtypes of PAAD and LIHC.

Discussion

It is well known that DDR genes have multifaceted roles in carcinogenesis. Elevated levels of DNA damage can play an initiating role in carcinogenesis, but cancer cells also require maintenance of DNA integrity to retain their elevated level of proliferation and to resist DNA damage from chemotherapeutic agents or radiotherapy[22–24]. Non-oncogene addiction is a principle whereby cancer cells become reliant on altered levels of unmutated proteins to promote their survival[25]. Inhibition of these genes represents a therapeutic strategy to treat cancer. An example of this is PARP inhibition of the BRCA pathway to treat breast cancer[26].

Gastrointestinal tumors are among the more lethal cancers due to their late diagnosis and lack of effective treatments[1]. In order to identify potential treatments for gastrointestinal (GI) cancers, an attempt was made to identify the presence of non-oncogene addiction of DDR genes in five GI tumor types.

Towards this goal, the RNA expression of 217 DDR genes was initially investigated in COAD, ESCA, LIHC, PAAD, and STAD, via GEPIA2. It was found that 11 DDR genes were commonly dysregulated across all these cancer types. These genes (MSH2, RAD51, FANCI, UBE2T, FEN1, EXO1, PRKDC, H2AX, CHEK1, SHFM1, PCNA) were each significantly overexpressed in each of the 5 GI cancer types. The level of overexpression (from Log_2FC 1.062 to 3.423) was consistent with the possibility of non-oncogenic addiction of these DDR genes in GI cancer.

In order to explore the biological significance of these DDR genes, it was investigated whether their overexpression correlated with patient survival, via GEPIA2. The hazard ratio (HR) of each of these 11 DDR genes was analyzed in each GI cancer type. Significant survival differences in DDR genes in PAAD and LIHC were observed, but not in the other cancer types. In PAAD, above-median tumor expression of each of the genes MSH2, RAD51, FANCI, UBE2T, and FEN1 was correlated with a high hazard ratio (HR ranging from 1.7 to 1.9). In LIHC, above-median tumor expression of each of the genes MSH2, RAD51, FANCI, UBE2T, FEN1, EXO1, PRKDC, H2AX, and CHEK1 was correlated with a high hazard ratio (HR ranging from 1.5 to 2.1). The significant survival reduction that was observed is consistent with the possibility that PAAD and LIHC tumors are non-oncogene addicted to these DDR genes. Therefore, inhibition of these genes represents a possible therapeutic

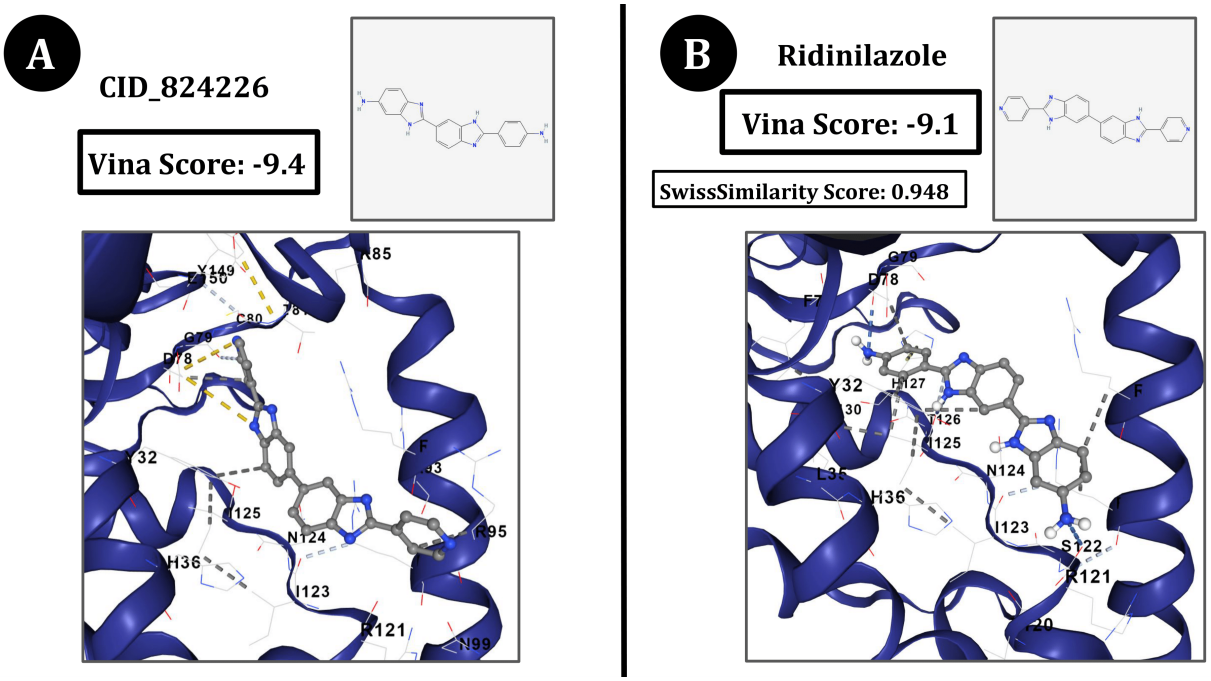
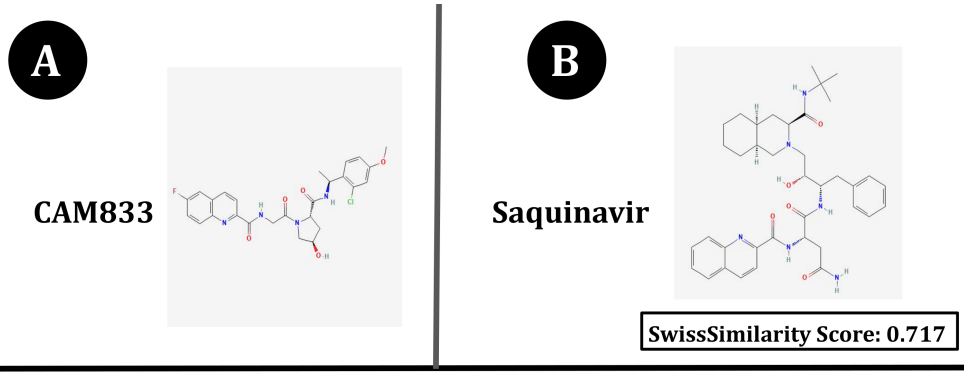


Fig. 5 (A) Vina score, chemical structure, and EXO1 binding site for the published inhibitor CID_824226 from CB-Dock2. (B) Vina score, chemical structure, and EXO1 binding site for the chemically similar inhibitor Ridinilazole from CB-Dock2. Amino acids at the EXO1-inhibitor binding pocket are shown.



C

	B02	RI-1	CAM833	Saquinavir
C1	-7.2	-6.5	-7.4	-7.9
C3	-7.4	-6	-8.3	-8.4
C5	-7.4	-6.2	-8.4	-7.7

Vina Score

Fig. 6 (A) Chemical structure of the published RAD51 inhibitor CAM833. (B) Chemical structure of Saquinavir, the similar drug to CAM833 identified from Swiss Similarity. (C) Binding affinity (Vina scores) of published inhibitors (B02, RI-1, and CAM833) compared to Saquinavir at the potential binding sites.

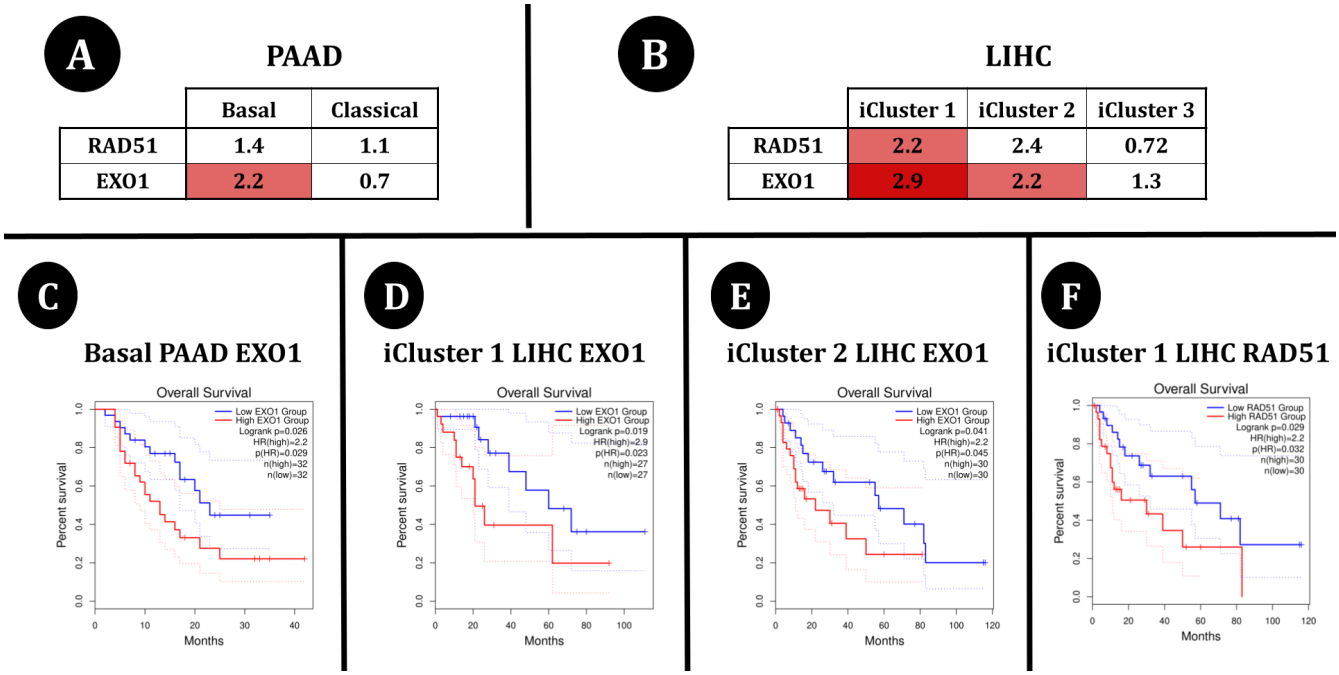


Fig. 7 (A) Hazard ratios of RAD51 and EXO1 in basal and classical pancreatic ductal adenocarcinoma (significant HR $>$ 1 and p $\leq$ 0.05 are shown in red). (B) Hazard ratios of RAD51 and EXO1 in iCluster 1, 2, and 3 subtypes in LIHC (significant HR $>$ 1 and p $\leq$ 0.05 are shown in red). (C) Kaplan-Meier chart for EXO1 in basal pancreatic ductal adenocarcinoma. (D) Kaplan-Meier chart for EXO1 in iCluster 1 LIHC. (E) Kaplan-Meier chart for EXO1 in iCluster 2 LIHC. (F) Kaplan-Meier chart for RAD51 in iCluster 1 LIHC.

approach to treat these cancers.

DDR genes exert their effects in several different DNA repair pathways, including homologous recombination, Fanconi Anemia, base-excision repair, mismatch-excision repair, and others. These 11 DDR genes that were found to be overexpressed in GI cancers were separated across 7 of these pathways. To identify whether a specific DDR pathway was critical in GI cancer survival, it was gauged whether the expression of these 11 DDR genes was correlated with each other, and whether any correlation had an impact on cancer survival.

Through GEPIA2 correlation analysis, individual gene pairs were compared to identify which ones were strongly correlated in expression (using the criteria Pearson's correlation coefficient R $\geq$ 0.6) in each of the above 9 LIHC genes and 5 PAAD genes. In addition, STRING analysis was also used to determine whether these 9 LIHC genes and 5 PAAD genes were physically or functionally connected. In LIHC, the correlation analysis revealed 7 genes (MSH2, RAD51, FANCI, UBE2T, FEN1, EXO1, and CHEK1) that were significantly correlated in expression with each other (Pearson's correlation coefficients ranging from 0.66 to 0.84), and subsequent STRING analysis reveal that these genes were in the same protein network (PPI Enrichment P-Value = 1.58e-14). In PAAD, the correlation analysis revealed 4 genes (RAD51, FANCI, UBE2T, and FEN1) that were significantly correlated in expression with each other (Pearson's correlation coefficients ranging from 0.64 to 0.81), and subsequent STRING analysis reveal that these genes were in the same protein network (PPI Enrichment P-Value = 1.95e-5). Subsequently, gene expression signatures for LIHC and PAAD were developed containing these correlated genes.

The signature for PAAD (RAD51, FANCI, UBE2T, and FEN1) had an HR of 1.5, while the signature for LIHC (MSH2, RAD51, FANCI, UBE2T, FEN1, EXO1, and CHEK1) had an HR of 1.8. Surprisingly, these signature HR values were within the range of the single gene HR values (ranging from 1.7 to 1.9 for PAAD, and 1.5 to 2.1 for PAAD). This indicated that there was no additive effect of these individual DDR genes in GI cancer survival for PAAD or LIHC, even though these genes represented several different DDR pathways. This suggests that inhibition of single DDR genes may be sufficient to produce a therapeutic effect and minimize off-target effects that may accompany the inhibition of multiple pathways.

To further delineate which of these several DDR genes may be the best targets for therapeutic intervention, their expression was analyzed in normal tissues via the Human Protein Atlas. Ideal therapeutic targets should have low expression in normal tissue, therefore minimizing any non-target side-effects[27]. Of the above 11 DDR genes, it was found that four genes in LIHC (RAD51, FANCI, EXO1, and CHEK1) and two genes in PAAD (RAD51 and FANCI) exhibited relatively low RNA or protein expression in normal tissues, supporting the possibility that these genes are good potential therapeutic targets. This conclusion

was confirmed by ShinyDepMap analysis, which showed that EXO1, RAD51, CHEK1, and FANCI, had efficacy scores of -0.478, -1.51, -1.666, and -0.582, respectively, suggesting that they promote cell growth in cancer cell lines.

RAD51 acts as a DNA binding protein in the homologous repair pathway[28]. RAD51 has been shown to be essential in tumor metabolism, metastasis and drug resistance. Therefore, inhibition of RAD51 is thought to be a possible therapeutic avenue to sensitize cancer cells to chemotherapeutic agents and other cancer treatments. It has been shown that RAD51 is upregulated in LIHC and inhibition of RAD51 reduces proliferation and enhances both apoptosis and DNA damage of LIHC cells[29]. It has also been shown that PAAD patients with higher levels of RAD51 expression had worse survival and that RAD51 promotes PAAD cell proliferation[30].

FANCI has been shown to be a predisposing gene for ovarian cancer and a prognostic biomarker for cervical cancer[31,32]. FANCI functions in DNA interstrand cross-link repair[33]. Additionally, FANCI inhibition has been shown to have a synergistic effect with PARP inhibitors, as FANCI inhibition has been shown to sensitize cancer cells to PARP inhibitors such as talazoparib[34]. FANCI was shown to have significantly higher expression in LIHC tissue and to be correlated with poor patient prognosis. FANCI was also found to facilitate tumor proliferation and metastasis as well as play a role in chemoresistance and immunosuppression in LIHC[35].

CHEK1 is a kinase activated by DNA damage that regulates DDR by activating checkpoints, delaying cell cycle progression, and enabling time for DNA damage repair. Higher CHEK1 levels have been shown to be associated with chemotherapy resistance and shorter overall patient survival[36]. CHEK1 in LIHC was found to have increased mRNA expression and to be correlated with overall worse survival outcomes[37]. Additionally, the knockdown of CHEK1 was shown to reduce the proliferation and aggression of LIHC cells.

EXO1 is a DDR enzyme that contributes to regulating cell cycle checkpoints, replication fork maintenance, and post-replicative DNA repair pathways[38]. EXO1 has been associated with breast, ovarian, lung, pancreatic, and gastric tract cancer. Overexpression of EXO1 has been linked with poor patient prognosis in a variety of cancers, including hepatocellular carcinoma[39]. EXO1 is also thought to be a potential drug target in cancer. In LIHC, EXO1 was also demonstrated to be an oncogene that when knocked down decreased proliferation and invasive capabilities[40].

Given the findings that RAD51, FANCI, CHEK1, and EXO1 are over-expressed in GI cancers, are associated with reduced survival, promote cell growth in cancer cell lines, have low expression in normal tissues, and have been found to have significant roles in other cancer types, these genes remained good targets for potential inhibitors. A strategy was devised to identify chemically similar molecules to known inhibitors of RAD51, FANCI, CHEK1, and EXO1, by screening a library of FDA-approved drugs. First, the literature was reviewed and identified existing experimental small molecule inhibitors known to suppress the activity of RAD51, EXO1, or CHEK1. Next, existing FDA-approved drugs that had a similar structure to these experimental small molecule inhibitors were identified through Swiss Similarity. Finally, the binding affinity of the FDA-approved drug with the DDR protein was evaluated through CB-Dock2.

For EXO1, CID_824226 was identified from the literature as an inhibitor, and Ridinilazole was found to be the FDA-Approved drug with the best similarity (Swiss Similarity Score, 0.948) that bound to this protein in silico with a high affinity (Vina score, -9.1 on CB-Dock2). For RAD51, RI-1, B02, and CAM833 were identified as the published inhibitors, and Saquinavir was found to be the FDA-Approved drug with the best similarity (Swiss Similarity Score, 0.717) that bound to this protein in silico with high affinity (Vina score of -8.4 on CB-Dock2). Finally, for CHEK1 7-hydroxy-staurosporine was identified as the published inhibitor, and Lestaurtinib as the repurposed FDA-approved drug with the best similarity (Swiss Similarity Score, 0.980) that bound to this protein in silico with high affinity (Vina score of -10.2 on CB-Dock2). To corroborate these results, specific chemical bonds between the drug and protein in question were also investigated using Arpeggio.

Saquinavir was found to have a slightly stronger binding affinity to RAD51, compared to all of the published inhibitors (RI-1, B02, and CAM833) at the binding sites C1 and C3, but had a slightly lower binding affinity at the C5 binding site. Saquinavir was originally approved by the FDA as a protease inhibitor, to treat HIV. Saquinavir was also shown to be a potential inhibitor of the main protease (3CLpro) of SARS-Cov-2[41]. Additionally, Saquinavir has been identified as a potential anti-cancer drug in various cancer types including cervical, prostate, and bladder cancer. These data suggest that Saquinavir may be a useful anti-cancer drug, given its strong binding to RAD51 in silico. However, further biological studies must be conducted. Saquinavir was shown to be generally well tolerated, with the most common side effect being mild gastrointestinal symptoms[42]. No published literature was found evaluating the use of Saquinavir in cancer patients.

Ridinilazole was found to have a comparable binding affinity as the published inhibitor (CID_824226) to EXO1, with Vina scores of -9.1 and -9.4, respectively. Ridinilazole is an FDA-approved antibiotic for the treatment of Clostridioides difficile[19]. However, there is no published literature, to our knowledge, associating Ridinilazole with anti-cancer properties. Therefore, Ridinilazole remains a promising target for further biological investigation as a potential novel EXO1 inhibitor and anti-cancer drug. 47.1% of patients treated with Ridinilazole in its Stage 3 trial had treatment-emergent adverse effects[43]. Of those, most were mild or moderate, with some (13.4%) of cases being severe.

Lestaurtinib was found to have a slightly stronger binding

affinity to CHEK1, as compared to the published inhibitor 7-hydroxy-staurosporine (Vina scores of -10.2 and -9.8 respectively). Further investigation with Arpeggio found this higher binding affinity could be due to an additional Carbon PI interaction between Lestaurtinib and Valine 23 of CHEK1. Considering that Lestaurtinib is a tyrosine kinase inhibitor, however, and that there is considerable cross reactivity with other tyrosine kinases, Lestaurtinib is not an optimal candidate as a specific CHEK1 inhibitor due to the potential for significant off-target effects[20].

Given the most promising DDR gene targets were RAD51 and EXO1, a subtype evaluation of these genes in LIHC and PAAD was undertaken. Evaluation of LIHC and PAAD subtypes revealed elevated hazard ratios for EXO1 and RAD51. EXO1 had an elevated hazard ratio in iCluster 1 (2.9) and iCluster 2 (2.2) of LIHC. RAD51 had an elevated hazard ratio in iCluster 1 (2.2). EXO1 had an elevated hazard ratio (2.2) in basal pancreatic ductal adenocarcinoma. This suggests that the potential anti-cancer properties of Saquinavir and Ridinilazole may be more potent in these specific cancer subtypes.

Limitations

These bioinformatics analyses are limited by the accuracy of the web tools used. The underlying tissue expression data from GTEx and TCGA utilized in this study are obtained from tissue blocks and not single cells, and therefore the expression of the genes investigated may not only reflect the cancer cells, but also may reflect the cells in the surrounding microenvironment. These databases are also subject to the experimental error wherein the underlying data was collected.

There also may be bias in the data collection used in these analyses as well. RNA sequencing data may only be collected frequently across patients with certain characteristics (e.g. late stage cancer) in a manner that could skew the datasets used. Additionally, smaller sample sizes of cancer patients remain an issue in bioinformatics analysis. In the GEPIA2 analyses across the 5 GI cancer types (COAD, ESCA, LIHC, PAAD, and STAD), sample sizes varied from 178 to 408 patients. The cancer subtype survival analysis has a notably smaller sample size (ranging from 54 to 64) relative to other analyses.

This study used GEPIA2's default differential expression analysis settings which may cause some genes relevant in cancer to be missed. However, strict search settings are required in this bioinformatics analysis in order to maintain the integrity of the results. These strict settings help counter inherent error that may be present with bioinformatics analysis. Additionally, strict criteria may be optimal in identifying the best drug targets. In this study we also excluded potential gene targets with elevated expression in normal tissues. This approach may also have discarded some good therapeutic targets, but the goal of these bioinformatics analyses is to identify the best potential targets, which should ideally have low expression in normal tissue to avoid off-target effects.

While this investigation proposes Saquinavir and Ridinilazole as already FDA-approved drugs with known properties, the pharmacokinetic and pharmacodynamic properties specifically in PAAD and LIHC are largely unknown. Therefore, further in vitro and in vivo studies are necessary to verify the computational drug docking results that were found.

There are several challenges in repurposing FDA approved drugs. Drugs might not have the same efficacy when rerouted to different targets. Consequently, higher doses may be needed, which could cause more substantial or even new adverse effects[44]. There also can be numerous regulatory, legal, and intellectual property challenges when repurposing drugs.

Conclusions

While treatment of cancer has often focused on oncogene targets, this study confirms the viability of cancer treatment through non-oncogenic addiction targets. In this study 11 DNA Damage Response genes were identified that displayed characteristics of non-oncogene addiction in gastrointestinal cancers. Subsequently, two FDA-approved drugs were identified through molecular modeling techniques, Saquinavir and Ridinilazole, which could potentially be repurposed to inhibit two of these promising gene targets, RAD51 and EXO1 respectively.

To corroborate these results, further biochemical and biological studies must be conducted to determine whether these drugs can inhibit RAD51 and EXO1, in vitro and in vivo. Specifically, an in vitro analysis of Saquinavir binding and inhibiting RAD51, and Ridinilazole binding and inhibiting EXO1 would validate the in-silico results. Additionally, the impact of RAD51 and EXO1 on LIHC, and RAD51 on PAAD can be evaluated using gene knockout techniques on cancer cell lines. Reduced expression of these DDR genes should decrease survival in these knockout cell lines, while the inhibitors Saquinavir and Ridinilazole should decrease survival in wild type cell lines. In vivo studies of tumor xenografts in mice could determine whether Saquinavir and Ridinilazole are able to reduce tumor growth and promote survival. Additional investigations could identify the efficacy of Saquinavir and Ridinilazole in conjunction with other cancer therapies such as chemotherapy, radiation therapy, and immunotherapy. If in vivo testing yields beneficial results, then a clinical trial can be undertaken for repurposing either Saquinavir or Ridinilazole as anti-cancer drugs.

If future studies determine that these drugs, Ridinilazole and Saquinavir, do inhibit RAD51 and EXO1 in vitro and in vivo, they could represent new treatments for liver and pancreatic cancer. However, this study already shows the bright future of bioinformatics and computational biology in the discovery of new cancer treatments.

References

1 M. Arnold, C. Abnet, R. Neale, J. Vignat, E. Giovannucci, K. McGlynn and F. Bray, *Global Burden of 5 Major Types of Gastrointestinal Cancer*, Epub 2020 Apr 2. PMID: 32247694; PMCID: PMC8630546.

2 D. Hanahan and R. Weinberg, *Hallmarks of cancer: the next generation*, PMID: 21376230.

3 C. Broustas and H. Lieberman, *DNA damage response genes and the development of cancer metastasis*, PMID: 24397478; PMCID: PMC4064942.

4 L. Bermúdez-Guzmán, *Pan-cancer analysis of non-oncogene addiction to DNA repair*, PMID: 34853396; PMCID: PMC8636604.

5 W. R. and L. M, *Human DNA repair genes [Internet*, https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html#Human, cited 2024 Aug 12]. Available from:.

6 Z. Tang, B. Kang, C. Li, T. Chen and Z. Zhang, *GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis*, PMID: 31114875; PMCID: PMC6602440.

7 D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. Gable, T. Fang, N. Doncheva, S. Pyysalo, P. Bork, L. Jensen and C. Mering, *The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest*, PMID: 36370105; PMCID: PMC9825434.

8 K. Shimada, J. Bachman, J. Muhlich and M. T. shinyDepMap, *a tool to identify targetable cancer genes and their functional connections from Cancer Dependency Map data*, PMID: 33554860; PMCID: PMC7924953.

9 P. Thul and C. Lindskog, *The human protein atlas: A spatial map of the human proteome*, PMID: 28940711; PMCID: PMC5734309.

10 J. Guo, H. Liu and J. Zheng, *SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets*, Epub 2015 Oct 29. PMID: 26516187; PMCID: PMC4702809.

11 V. Zoete, A. Daina, C. Bovigny and O. Michielin, *SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening*, PMID: 27391578.

12 E. Sayers, E. Bolton, C. Brister, JR, C. K, C. J, C. DC, F. R, K. K, K. C, M. S, M.-B. T, L. A, L. C, L. S, T.-N. Z, M. F, P. T, S. L, T. Y, W. T, W. J, T. R, P. BW, S. KD and S.T., *Database resources of the national center for biotechnology information*, PMID: 34850941; PMCID: PMC8728269.

13 Y. Liu, X. Yang, J. Gan, S. Chen, Z. Xiao and Y. Cao, *CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting*, PMID: 35609983; PMCID: PMC9252749.

14 P. Rose, A. Prlić, A. Altunkaya, C. Bi, A. Bradley, C. Christie, L. Costanzo, J. Duarte, S. Dutta, Z. Feng, R. Green, D. Goodsell, B. Hudson, T. Kalro, R. Lowe, E. Peisach, C. Randle, A. Rose, C. Shao, Y. Tao, Y. Valasatava, M. Voigt, J. Westbrook, J. Woo, H. Yang, J. Young, C. Zardecki, H. Berman and S. Burley, *The RCSB protein data bank: integrative view of protein, gene and 3D structural information*, Epub 2016 Oct 27. PMID: 27794042; PMCID: PMC5210513.

15 *The PyMOL Molecular Graphics System, Version 1.2r3pre*, https://pymol.org/.

16 H. Jubb, A. Higueruelo, B. Ochoa-Montaño, W. Pitt, D. Ascher and T. Blundell, *Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures*, Epub 2016 Dec 10. PMID: 27964945; PMCID: PMC5282402.

17 S. Spruance, J. Reid, M. Grace and M. Samore, *Hazard ratio in clinical trials*, PMID: 15273082; PMCID: PMC478551.

18 M. Kircher, S. Sawyer and M. Meyer, *Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform*, PMID: 22021376; PMCID: PMC3245947.

19 A. Gonzales-Luna, T. Carlson and K. Garey, *Emerging Options for the Prevention and Management of Clostridioides difficile Infection*, Epub 2023 Jan 16. PMID: 36645620; PMCID: PMC9841950.

20 E. Busby, D. Leistritz, R. Abraham, L. Karnitz and J. Sarkaria, *The radiosensitizing agent 7-hydroxystaurosporine (UCN-01) inhibits the DNA damage checkpoint kinase hChk1*.

21 R. Shen, A. Olshen and M. Ladanyi, *Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis*, PMID: 19759197; PMCID: PMC2800366.

22 P. Bouwman and J. Jonkers, *The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance*.

23 E. Abad, D. Graifer and A. Lyakhovich, *DNA damage response and resistance of cancer stem cells*, https://doi.org/10.1016/j.canlet.2020.01.008., ISSN 0304-3835,.

24 F. Sadoughi, L. Mirsafaei, P. M. Dana, J. Hallajzadeh, Z. Asemi, M. A. Mansournia, M. Montazer, M. Hosseinpour and B. Yousefi, *The role of DNA damage response in chemo- and radio-resistance of cancer cells: Can DDR inhibitors sole the problem?*

25 J. Luo, S. N. L and S. Elledge, *Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction*.

26 J. Brown, B. O'Carrigan, S. Jackson and T. Yap, *Targeting DNA Repair in Cancer: Beyond PARP Inhibitors*, Epub 2016 Dec 21. PMID: 28003236; PMCID: PMC5300099.

27 Duffy, M. Verbanck, A. Dobbyn, H. Won, J. Rein, I. Forrest, G. Nadkarni, G. Rocheleau and R. Do, *Tissue-specific genetic features inform prediction of drug side effects in clinical trials*, PMID: 32917698; PMCID: PMC11206454.

28 Z. Wang, R. Jia, L. Wang, Q. Yang, X. Hu, Q. Fu, X. Zhang, W. Li and Y. Ren, *The Emerging Roles of Rad51 in Cancer and Its Potential as a Therapeutic Target*, PMID: 35875146; PMCID: PMC9300834.

29 M. Pan, Y. Sha, J. Qiu, Y. Chen, L. Liu, M. Luo, A. Huang and J. Xia, *RAD51 Inhibition Shows Antitumor Activity in Hepatocellular Carcinoma*.

30 X. Zhang, N. Ma, W. Yao, S. Li and Z. Ren, *RAD51 is a potential marker for prognosis and regulates cell proliferation in pancreatic cancer*, PMID: 31889908; PMCID: PMC6935217.

31 C. Fierheller, W. Alenezi, C. Serruya, T. Revil, S. Amuzu, K. Bedard, D. Subramanian, E. Fewings, J. Bruce, S. Prokopec, L. Bouchard, D. Provencher, W. Foulkes, Z. El Haffaf, A. Mes-Masson, M. Tischkowitz, I. Campbell, T. Pugh, C. Greenwood, J. Ragoussis and P. Tonin, *Molecular Genetic Characteristics of FANCI, a Proposed New Ovarian Cancer Predisposing Gene*, PMID: 36833203; PMCID: PMC9956348.

32 X. Liu, X. Liu and X. Han, *FANCI may serve as a prognostic biomarker for cervical cancer: A systematic review and meta-analysis*, PMID: 34941027; PMCID: PMC8702066.

33 S. Sondalle, S. Longerich, L. Ogawa, P. Sung and S. Baserga, *Fanconi anemia protein FANCI functions in ribosome biogenesis*, Epub 2019 Jan 28. PMID: 30692263; PMCID: PMC6377447.

34 Y. Huang, M. Sang, P. Xi, R. Xu, M. Cai, Z. Wang, J. Zhao, Y. Li, J. Wei and Q. Ding, *FANCI Inhibition Induces PARP1 Redistribution to Enhance the Efficacy of PARP Inhibitors in Breast Cancer*, doi: 10.1158/0008-5472.CAN-23-2738. Epub ahead of print. PMID: 39037758.

35 Y. Hou, J. Li, A. Yu, K. Deng, J. Chen, Z. Wang, L. Huang, S. Ma and X. Dai, *FANCI is Associated with Poor Prognosis and Immune Infiltration in Liver Hepatocellular Carcinoma*, PMID: 37324186; PMCID: PMC10266051.

36 F. Neizer-Ashun and R. Bhattacharya, *Reality CHEK: Understanding the biology and clinical potential of CHK1*, Epub 2020 Sep 28. PMID: 32991949.

37 E. Bai, M. Dong, X. Lin, D. Sun and L. Dong, *Expressional and functional characteristics of checkpoint kinase 1 as a prognostic biomarker in hepatocellular carcinoma*, PMID: 36644193; PMCID: PMC9834594.

38 G. Keijzers, D. Bakula, M. Petr, N. Madsen, A. Teklu, G. Mkrtchyan, B. Osborne and M. Scheibye-Knudsen, *Human Exonuclease 1 (EXO1) Regulatory Functions in DNA Replication with Putative Roles in Cancer*, PMID: 30585186; PMCID: PMC6337416.

39 Y. Dai, Z. Tang, Z. Yang, L. Zhang, Q. Deng, X. Zhang, Y. Yu, X. Liu and J. Zhu, *EXO1 overexpression is associated with poor prognosis of hepatocellular carcinoma patients*, Epub 2018 Oct 20. PMID: 30328366; PMCID: PMC6237436.

40 G. Yang, K. Dong, Z. Zhang, E. Zhang, B. Liang, X. Chen and Z. Huang, *EXO1 Plays a Carcinogenic Role in Hepatocellular Carcinoma and is related to the regulation of FOXP3*, PMID: 32626539; PMCID: PMC7330697.

41 M. Pereira and N. Vale, *Saquinavir: From HIV to COVID-19 and Cancer Treatment*, PMID: 35883499; PMCID: PMC9313067.

42 S. Vella and F. M. Saquinavir, *Clinical pharmacology and efficacy*, PMID: 9533981.

43 P. Okhuysen, M. Ramesh, T. Louie, N. Kiknadze, J. Torre-Cisneros, C. Oliveira, C. Steenkiste, A. Stychneuskaya, K. Garey, J. Garcia-Diaz, J. Li, E. Duperchy, B. Chang, J. Sukbuntherng, J. Montoya, L. Styles, F. Clow, D. James, E. Dubberke, W. M. Randomized and Double-Blind, *Phase 3 Safety and Efficacy Study of Ridinilazole Versus Vancomycin for Treatment of Clostridioides difficile Infection: Clinical Outcomes With Microbiome and Metabolome Correlates of Response*, PMID: 38305378; PMCID: PMC11175683.

44 Y. Xia, M. Sun and H. Huang, *Drug repurposing for cancer therapy*.