Exploiting Synthetic Lethal Network for Precision Treatment of Clear Cell Renal Cell Carcinoma

Zhicheng Liu
Huazhong University of Science and Technology Tongji Medical College https://orcid.org/0000-0002-5950-409X

Dongxu Lin
Tongji Medical College of Huazhong University of Science and Technology: Huazhong University of Science and Technology Tongji Medical College

Linmeng Zhang
Shanghai Cancer Institute

Chen Yang
Shanghai Cancer Institute

Bin Guo
Tongji Medical College of Huazhong University of Science and Technology: Huazhong University of Science and Technology Tongji Medical College

Feng Xia
Tongji Medical College of Huazhong University of Science and Technology: Huazhong University of Science and Technology Tongji Medical College

Yan Li
Zhongshan University: Sun Yat-Sen University

Danyang Chen
Tongji Medical College of Huazhong University of Science and Technology: Huazhong University of Science and Technology Tongji Medical College

Cun Wang
Shanghai Cancer Institute

Zhenyu Xiao (xiaozhenyu58@163.com)
Tongji Medical College of Huazhong University of Science and Technology: Huazhong University of Science and Technology Tongji Medical College

Zhong Chen
Tongji Medical College of Huazhong University of Science and Technology: Huazhong University of Science and Technology Tongji Medical College

Chao Leng
Tongji Medical College of Huazhong University of Science and Technology: Huazhong University of Science and Technology Tongji Medical College

Research Article

Keywords: renal cancer, synthetic lethality, precision medicine, drug repurposing, driver mutation

Posted Date: December 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-958499/v1

Abstract

Background

The emerging of targeted therapies has revolutionized the treatment modalities of advanced clear cell renal cell carcinoma (ccRCC) over the past fifteen years. However, lack of personalized treatment limits the development of effective clinical guidelines and improvement of patient prognosis. In this study, large-scale genomic profiles of ccRCC cohorts were exploited for conducting an integrative analysis.

Method

Based on synthetic lethality (SL), a concept that simultaneous losses of two genes cause cell death while a single loss does not, we sought to develop a computational pipeline to infer potential SL partners of ccRCC. Drug response prediction were received from three pharmacological databases to select agents which are likely to be effective in precisely treating patients with target gene mutation.

Results

We developed a credible method to identify SL pairs and determined a list of 72 candidate pairs which might be utilized to selectively eliminate tumors with genetic aberrations through SL partners of specific mutations. Further analysis identified *BRD4* and *PRKDC* as novel medicine targets for patients with *BAP1* mutations. After mapping these target genes to comprehensive drug datasets, two agents (BI-2536 and PI-103) were found to have considerable therapeutic potential in *BAP1* mutant tumors.

Conclusion

Overall, our findings provide insight into the overview of ccRCC mutation patterns and offer novel opportunities for improving individualized cancer treatment.

Introduction

Renal cell carcinoma (RCC) is one of the most lethal malignancies in the genitourinary system. A recent study showed that there were 431,288 (2.2%) new cases and 179,368 (1.8%) deaths of RCC in 2020 [1]. Approximately 70% of renal cancer patients present with localized stage, suggesting a possible complete excision of tumor by radical nephrectomy [2, 3]. Clear cell renal cell carcinoma (ccRCC) is the most prevalent subtype, accounting for more than 70% of all RCC [4]. Although most ccRCC can accept effective treatment, including surgery or ablation when early diagnosed, the rate of distant metastasis is up to a third after treatment [5]. Considering the poor prognosis of ccRCC patients, more efforts are required to develop optimal adjuvant or targeted therapies benefiting these patients.

With the rapid development of genome sequencing, tremendous genomic information of carcinoma was identified, which has proved the significant role of driver mutation (DM) in the occurrence and development of renal cancer [6]. And genetically targeted drugs have been successfully applied to remedy patients with mutated genes. VEGFR inhibitor sunitinib and mTOR signaling inhibitor everolimus are two representative agents for renal cancer therapy. However, there are still many patients who suffer from tumor recurrence due to drug resistance. The synthetic lethality (SL) strategy, providing a promising approach to curing these patients, has been successfully applied to searching tailored and effective anticancer compounds, such as Poly (ADP-ribose) polymerase (PARP) inhibitors olaparib. Briefly, the simultaneous mutation of a specific gene pair will cause tumor cell death, and the functional loss of either one brings little effect on cell survival. Despite many challenges to pharmacologically rescue the function of mutation genes such as von Hippel-Lindau (*VHL*) and BRCA1 associated protein 1 (*BAP1*), the application of drugs targeting the second-site of SL pairs was taken as an alternative method in treating patients affected by gene mutation. Harnessing this concept, current studies have focused on identifying the SL gene pairs associated with *VHL* - hypoxia inducible factor (*HIF*) signaling [7–9]. In order to find more SL gene pairs with therapeutic potential, it is necessary to expand the scope of screening.

Previous research has proposed various algorithms to identify SL gene-pairs, such as *DAISY* [9] and *MiSL* [10]. Nevertheless, such procedures mainly used non-specific inference for pan-cancer analysis, which could be unsuitable for renal cancer patients with specific mutation patterns. In the present study, we conducted a comprehensive literature retrieval to search for publicly available data of ccRCC. Then a novel strategy of SL interaction analysis was applied to identify potential SL gene-partners of driver genes in ccRCC. The paired genes with therapeutic implications were identified after filtering out the candidate SL pairs, and compounds collected from multiply drug databases were matched to them to identify potential therapeutic therapies for tumor patients. Generally, our findings may provide comprehensive insight into the mutation pattern of ccRCC, and new opportunities for exploring highly specific therapeutic strategies against renal tumors.

Materials And Methods

RNA-sequencing cohorts

In total, five RNA-sequencing cohorts of ccRCC, including the Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) cohort [11], renal cell cancer-EU (RECA-EU), CheckMate 009 (CM-009) [12], CheckMate 010 (CM-010) [13] and CheckMate 025 (CM-025) [14] were used in this study. Of these, gene-expression, mutation profiles and full clinical annotations of TCGA-KIRC, RECA-EU were obtained from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/repository) and the International Cancer Genome Consortium (ICGC) portal (https://dcc.icgc.org/). The relevant information about CM cohorts was achieved from the supplementary files of three prospective clinical trials which comprised of ccRCC patients treated with anti-PD-1 antibody immunotherapy [15]. All expression data (raw counts) of RNA-sequencing datasets mentioned above were transformed into transcripts per million (TPM) values and these RNA-seq cohorts were integrated into one combined metadata. The ComBat algorithm of *SVA* R

package [16] was applied to correct batch effects from non-biological technical biases to ensure comparability between different cohorts (Additional file 1: Fig. S1A). The single nucleotide variants (SNVs) and small insertions/deletions (INDELs) of mutation data were saved for further analysis, while copy number variants (CNVs) profiles were not included due to the data limitation. In order to evaluate the effect of mutation on gene expression, the expression data and functional mutations were involved in this study. Notably, functional mutations, including frameshift and nonsense mutations, were defined as alternations that the resulting proteins usually affected normal physiological functions of cells. The non-functional mutations, including silent mutations (synonymous mutations) were excluded and samples with no functional mutations or fewer than 10 mutations in gene panels were considered as outliers and discarded from downstream analyses. Genes with duplicated mutations were merged to keep only one record.

Microarray cohorts

The expression data, somatic mutations data and clinical information of E-MTAB-1980 (including 101 ccRCC samples based on GPL13497), E-MTAB-3218 (including 114 ccRCC samples based on GPL13667), E-MTAB-3267 (including 59 ccRCC samples based on GPL6244) were acquired from the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/). Then background adjustment and quantile normalization were performed on these raw expression files from Affymetrix and Agilent by using the robust multiarray average (RMA) method located in R package *Affy* [17]. For GSE29609 cohort (including 39 ccRCC samples based on GPL1708), the expression data and detailed clinical information were collected from the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and the raw expression data were also normalized by the RMA method. These microarray cohorts were merged into one combined cohort with batch effect removal using ComBat function (Additional file 1: Fig. S1B). Additionally, the mutation annotation information of GSE29609 cohort is unavailable.

Cancer cell line data

Gene expression profiles and somatic mutation data of human cancer cell lines (CCLs) were downloaded from the Cancer Cell Line Encyclopedia (CCLE) project (https://portals.broadinstitute.org/ccle/) and Genomics of Drug Sensitivity in Cancer (GDSC) project (https://www.cancerrxgene.org/). The experimental information of different drug responses against CCLs was achieved from the Cancer Therapeutics Response Portal (CTRP v2.0, released October 2015, https://portals.broadinstitute.org/ctrp), PRISM Repurposing dataset (19Q4, released December 2019, https://depmap.org/portal/prism/) and GDSC 1&2 datasets (Release 8.2, release February 2020, https://www.cancerrxgene.org/downloads/bulk_download), respectively. Of these medicine databases, PRISM contained the drug sensitivity data of 1,448 compounds against 499 CCLs, CTRP provided the drug sensitivity data of 545 compounds against 907 CCLs, and GDSC included the drug sensitivity data of 518 compounds against 988 CCLs. And the area under curve (AUC) values of dose-response acquired from these three datasets were used as evaluation indicators of drug sensitivity, which lower AUC value suggests higher response probability to therapy treatment. Compounds with missing AUC values across more than 20% of the CCLs were excluded firstly, and the rest of compounds containing incomplete data

were imputed using the K-nearest neighbors (KNN) method [18] located in R package *Impute*. Notably, expression profiles and molecular data of CCLs were downloaded from the same CCLE Project, and were used for subsequent PRISM and CTRP analyses. In order to investigate the cancer survival-essential genes, the genome-wide gene dependency scores, including CERES scores from clustered regularly interspaced short palindromic repeats (CRISPR) knockout screens [19] and DEMETER scores from RNA interference (RNAi) screens [20], were achieved from the Cancer Dependency Map (DepMap) portal (https://depmap.org/portal/download/), which lower CERES or DEMETER scores denote that relevant genes are more likely to be essential in cell survival and proliferation of CCLs.

BAP1 mutation prediction

Due to missing mutation data of part samples in E-MTAB-3267, E-MTAB-3218 and GSE29609, elastic net (EN)-based prediction model, a generalized linear model in the R package *glmnet* [21], was utilized to forecast *BAP1* mutation status. The RNA-seq metadata mentioned above were then used as training cohort to construct the prediction model, and samples with mutation annotations in E-MTAB-1980 were considered as external validation for evaluating the performance of *BAP1* prediction model. To select significant genes which were taken as input into EN model (abs (Log2FC) > 1.5 & adjust $P < 0.05$), differential expression analysis between the *BAP1* mutant and WT samples from the training cohort was performed using the R package *limma* [22]. Additionally, the leave-one-study-out cross-validation was performed to evaluate the accuracy of EN model. Specifically, after splitting a dataset into a training set and a testing set, and using all but one observation as part of the training set, the prediction model was built using data from the training set. Lastly, this process was repeated *n* times (where *n* is the total number of observations in the dataset), leaving out a different observation from the training set each time, which meant that it provided a much less biased measure of test mean squared error compared to other cross-validation methods. Notably, the penalty was set as 0.9 in fitting a generalized linear model. The predictive performance of the EN model in training and validation cohorts was evaluated using receiver operating characteristic (ROC) curve via the R package *pROC* [23].

Detection of cancer driver mutations

To discern likely DMs regulating gene network of tumor expression from thousands of mutations, the *DriverNet* algorithm [24] was applied in this study, which could evaluate the DM probability through integrating genome and expression data. Accordingly, a mutation matrix, a corresponding expression matrix and an influence graph were taken as input documents of *DriverNet*. In this analysis, the influence graph was derived from the Reactome Functional interactions [25], an updated protein functional interaction network (Version 2020). Notably, the results of *DriverNet* indicated the probabilities whether imported mutations belong to DMs, and genes with P value < 0.05 were deemed statistically significant. To make our prediction more reliable, we compiled a comprehensive list of cancer-associated driver genes which have been validated from prior studies and made a comparison between our prediction and previous results. These same DMs were saved for constructing the network between DMs and druggable genes (DGs) subsequently.

Collection of drug-target interactions

The medicine information about drug-target was acquired from the Drug Repurposing Hub [26] and DrugBank [27], respectively. The Drug Repurposing Hub (released March 2020, https://clue.io/repurposing#download-data) contained 6798 unique compounds and 2,183 targeted genes, and DrugBank (Version 5.1.8, released January 2021, https://go.drugbank.com/releases/latest) comprised 7,540 compounds and corresponding 3,976 targeted genes. Then two drug data were merged into one meta-drug set, and a total of 11,875 compounds and 4,465 DGs were identified after removing duplicated medicine information. In order to identify genes with potential therapeutic implications, DGs were utilized to construct DM-DG-drug network.

Mutual exclusivity analysis

Under the SL hypothesis, no somatic alteration happens on both genes of candidate partners in ccRCC simultaneously. Based on the somatic mutation data of 1,211 patients, the analysis was performed by using the *DISCOVER* R package to determine significant mutual exclusivity [28]. Gene pairs with adjusted *P* value < 0.1 were considered statistically significant.

Connectivity map analysis

To identify potentially therapeutic compounds, connectivity map (CMap) analysis (https://clue.io/) was used for searching compounds of which gene expression patterns were opposite to the *BAP1* mutant expression pattern. Differential analysis between *BAP1* mutant and WT samples was performed to select 150 up-regulated and down-regulated genes with the most significant fold changes respectively. Through the CMap analysis, the standardized connectivity score for each perturbation was calculated, which ranges from -100 to 100. Compounds with the CMap score < -80 were considered to have a potential therapeutic effect for ccRCC.

Identification of ccRCC subclasses

Network-based stratification (NBS) was performed to identify subclasses of ccRCC via Python package *pyNBS* [29], which divides tumor samples with available somatic mutation profiles into molecularly and clinically relevant subtypes on the basis of the mutation characters of the combined RNA-seq cohort [30]. Through integrating a high-quality cancer reference network from the recent study [29] and a mutation matrix of driver genes, we acquired the resulting data which contained the clustering information and corresponding consensus matrix from NBS. To evaluate the robustness of clusters *k* ranging from 2 to 5, the cophenetic correlation coefficient was calculated using the R package *NMF* [31] and the value of *k* that resulted in the maximum cophenetic correlation coefficient was considered as the optimal number of clusters. In addition, the nearest template prediction (NTP) analysis was conducted via R package *CMScaller*, which could predict the previously published RCC classifications based on the provided subclass signatures [32].

Functional similarity analysis

In this study, *GOSemSim*, an R package for measuring semantic similarity among GO terms and gene products [33], was utilized to estimate the similarity of molecular function (MF) and cellular component (CC) among different genes. Gene pairs achieved from DM-DG network above were used to measure the functional similarity score (FSS), which was calculated based on the semantic similarity in MF (SsMF) and CC (SsCC), as following formula:

$$Functional\ Similarity\ Score = \sqrt{SsMF * SsCC}$$

Notably, gene pairs with FSS > 0.45 were considered to have high functional correlations and were used for further analysis.

Rank aggregation analysis

To obtain a consistent result across multiple sources, *rank aggregation* algorithm, an order statistics-based method located in R package *RobustRankAggreg* proposed by Kolde et al [34], was applied in this study, of which the result (*P* value) indicates whether the ranking of a particular gene pair is statistically significant. In this analysis, we chose the order statistics method proposed by Stuart et al [35] by assigning the corresponding parameter to '*the Stuart*' and defined the rank aggregation score (RAS) as follows:

$$Rank\ \ Aggregation\ \ Score = \ -log2(P\ \text{value})$$

The ranking of candidate gene pairs was determined by the RAS, and a higher RAS denoted a more concordant ranking.

Predicting drug response in clinical samples

Three large pharmacogenomic datasets, including CTRP, PRISM and GDSC, contained massive drug screening and gene expression data across hundreds of cancer cell lines. Previous studies have demonstrated that drug response in clinical samples can be predicted using data from in vitro cell line experiments [36]. To perform drug response prediction, we intended to test different machine learning methods, including support vector machine, random forest and multivariate linear regression, based on the actual drug sensitivity and molecular data. In this study, the ridge regression model that exhibited great and precise performance in the previous research [37] was utilized for transcriptome data-based drug response prediction using the R package *pRRophetic*. Through exploiting the expression and drug response data of solid CCLs from CCLE and GDSC projects (excluding hematopoietic and lymphoid tissue-derived CCLs), this predictive model was trained with a satisfied predictive accuracy evaluated by default 10-fold cross-validation and then applied to calculate different drug response across clinical samples. These compounds with positive response calculated by this model were matched to their DGs for subsequent construction of the DM-DG-drug network.

Enrichment analysis

We performed gene set variation analysis (GSVA) using the R package *GSVA* based on the hallmark definitions (h.all.v7.4.symbols) extracted from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/) [38] to explore the differential expression of certain pathway or signature between *BAP1* and WT patients [39]. Notably, the resulting *P value* from the hypergeometric test was adjusted for multiple comparison testing and *P adjust* < 0.05 was considered significant.

Statistical analysis

All the statistical tests and graphical visualization were conducted utilizing R statistical software, version 4.0.5 (https://cran.r-project.org/). Student's t-test or Wilcoxon rank-sum test was applied for comparison of two groups with or without normally distributed variables, respectively. Similarly, correlation between two continuous variables was measured by either Pearson's r correlation (measure of linear relationship between two continuous variables) or Spearman's rank-order correlation (nonparametric measure of statistical dependence between two variables). Contingency table variables were analyzed by Fisher's exact tests. The Kaplan–Meier method was applied to perform survival analysis and the statistical significance of differences was determined using the log-rank (Mantel-Cox) test. The hazard ratios (HR) were calculated using the univariate Cox proportional hazards regression model located in R package *survival*. The Benjamini-Hochberg method was utilized to adjust *P* value of multiple testing in those analyses with more than 20 comparisons. *P* value < 0.05 was considered statistically significant for all computational analysis unless otherwise stated.

Result

Overview of the SL interaction analysis

A total of 1,174 ccRCC transcriptome profiles together with clinical information were collected from numerous publicly available cohorts, including TCGA-KIRC, RECA-EU, CM-009, CM-010, CM-025, E-MTAB-1980, E-MTAB-3218, E-MTAB-3267 and GSE29609. Of these patients, 928 were from RNA sequencing data, which were used for further SL framework construction, and 246 were from microarray data, which were considered as external validation for evaluating the result of DM-DG-drug network. A Schematic diagram of the procedures of SL inference and the overall study design was presented in Fig. 1.

The SL interaction analysis is a computational pipeline for identifying candidate SL interactions drawing on the experiences from several previous researches, such as *DAISY* [9], *MiSL* [10], *SELECT* [40]. This analysis consists of four statistical inference procedures:

1. Differential gene expression: The procedure exploited gene expression and somatic alternations of the inputting tumor samples to discover potential SL gene pairs under the assumption that carcinoma cells may increase the expression of its SL partners as a compensatory mechanism when a driver gene loses its function due to the mutation. Differential expression analysis was conducted using Wilcoxon rank-sum test between samples with and without DMs and only target genes with higher expressions in mutated samples were saved as potential SL partners of corresponding DMs.

2. Pairwise gene co-expression: The procedure tended to select gene pairs which could have similar functions of cell metabolism and growth, and be likely to co-express in para-carcinoma normal tissues on the notion that there is often an intensive relationship between both genes of SL pair. Gene pairs presenting significant correlations (Spearman correlation coefficient > 0.1 and *P adjust* < 0.05) were considered as SL candidate pairs.
3. Functional similarity: The procedure aimed to filter out gene pairs having high semantic similarity, motivated by the assumption that SL partners tend to engage in closely related biological processes. And accordingly, their locations in Gene Ontology (GO) topological structure should be close. FSS which was defined as the geometric mean of semantic similarities of MF and CC ranges from 0 to 1. And FSS $\leq$ 0.45 between gene pairs were considered to have no significant functional similarity and thus excluded from the candidate SL pairs.
4. Mutual exclusivity: The procedure selected those gene pairs which the incidence of simultaneous mutation was significantly lower than common gene pairs, based on the concept that simultaneously mutating two genes in an SL pair would affect the cellular process and cause tumor cell death. The gene pairs with *P adjust* < 0.15 were considered as potential SL pairs.

Those candidate pairs passing all the four procedures composed the final output set of candidate SL pairs and were subsequently used for constructing the DM-DG-drug network.

Detection of driver genes in ccRCC

The current consensus viewpoint on tumorigenesis and tumor progression was that only a few mutational events affecting driver genes were determined to be the origin of malignancy, which confers selective growth advantage to the tumor cell. Compared with traditional chemicals, small molecule compounds targeting DM have the advantageous property of avoiding impairment of normal tissue, and thus screens on these DMs are more likely to identify clinically significant targets. In this study, to identify candidate drivers, the DriverNet algorithm was applied in the most comprehensive metadata set of ccRCC currently, which contained 610 patients from five clinical cohorts with both expression and mutation data available (Fig. 2A). A total of 36 candidate genes had been yielded with *P adjust* < 0.1 and mutation frequency beyond mean (Additional file 2: Table S1). Notably, due to the limitations of the influence graph derived from the Reactome Functional interactions, *SETD2* which has been confirmed as driver genes in previous studies [11, 41, 42] was added in our prediction model to achieve more reliable results. Of these genes, 25 genes (67.6%) which demonstrated the reliability of our prediction have been reported by at least one previous research and were then taken as robust drivers of ccRCC for subsequent analysis.

To explore the clinical implications of DMs in ccRCC, the NBS algorithm was applied to stratify patients into different subtypes utilizing their mutation profiles. According to the resulting cophenetic correlation coefficients, 610 patients were assigned into two groups (Additional file 1: Fig. S2A and S2B). The result indicated that each group had distinguishing mutation features (Fig. 2B). NBS2 contained a higher proportion of common DMs, including *VHL*, *PBRM1* and *SETD2*, while NBS1 exhibited a high frequency of *BAP1* and increased mutational burden (Fig. 2C and Additional file 1: S1C). In addition, we analyzed the

relationship between the NBS classification and the previously reported RCC molecular subgroups, including Rini's (Low-High recurrence score group) [43], Brooks' (ccA-ccB group) [44] and Motzer's (Poor-Favorable risk group) [45]. NBS1 was found to be positively associated with Rini's High recurrence score group (P = 0.1713), Brooks' CCB group (P = 0.046) and Motzer 's Poor risk group (P = 0.0416), while NBS2 exhibited opposite patterns (Additional file 2: Table S2). Next, the correlation between NBS classification and clinical characteristics, containing the clinical stage, pathological stage and survival time, was investigated in the combined cohort. There was a significant difference in survival outcome between NBS groups, which NBS2 exhibited a better prognosis than NBS1(P = 0.0021) (Fig. 2D). However, other clinical characteristics were found to be weakly correlated with NBS classification (Additional file 2: Table S2). Taking together, the NBS classification provided novel insight into the DM-based clinical subclass of ccRCC patients and enhanced our understanding of the crucial role which driver genes played in tumorigenesis and development.

Selection of druggable genes

The SL candidates of DMs can be achieved leveraging our computational pipeline, while there encounters another problem that not all identified partners of DMs could be targeted when performing genome-wide scanning for potential SL partners. Therefore, to infer statistically candidate SL partners which could be targeted with conventional chemical agents, a list of 4,465 DGs was compiled from current public pharmacological databases and considered as the input of SL analysis. Of these DGs, only 1,981 targets were used for constructing DM-DG network due to low expression of some DGs after removing batch effects.

Inference of driver mutation-druggable gene interactions

Based on the above mentioned 25 DMs and 1,981 DGs, SL interaction analysis was conducted to infer DM-DG pairs which meet the corresponding criteria. In total, 72 DM-DG pairs (containing 69 unique drug targets) passed all the screening procedures and thus composed SL candidates for ccRCC (Fig. 3A). Additionally, rank aggregation analysis was performed to integrate the results of each procedure in SL interaction in order to obtain a robust ranking of the 72 DM-DG pairs. Accordingly, the ranks of candidate pairs were ordered based on FS scores (functional similarity), fold change values (differential expression), correlation coefficients (pairwise co-expression), and *P adjust* (mutual exclusivity) respectively. Then, the Stuart method was applied to integrate all the rankings and calculated the RAS of each DM-DG pair (Additional file 2: Table S3).

To validate whether the DM-DG pair exhibits SL interaction, we performed univariate survival analysis between DG expression in patients with specific DM and progression-free survival (PFS) using Cox proportional hazards regression models. Part significant DGs were found to be associated with shorter recurrence time (HR>1) among patients with relevant DMs (Additional file 2: Table S4). Additionally, Kaplan-Meier analysis was conducted to reveal the clinical relationship between PFS and the status of DG in patients with corresponding DM. Specifically, we mainly defined the functional status of one gene by dividing expression data into active (>median) and inactive groups (<median) for the lacking of

aberration situation of DGs (Fig. 3B and 3C). As depicted from the figure, *BRD4* and *TYK2* inactive groups had significant survival advantage in ccRCC patients with *BAP1* and *VHL* mutations compared with active groups, respectively. These survival data-based analyses demonstrated the clinical phenotypes of these 72 DM-DG pairs could be well compatible with their roles as SL candidates.

Estimation of drug response in clinical samples

Three pharmacogenomic datasets described in the Materials and Methods section, containing drug sensitivity data and gene expression profiles of multiple CCLs, were utilized to construct the drug prediction model. Notably, chemical compounds with NAs in more than 20% of the samples and CCLs derived from hematopoietic and lymphoid tissue were excluded to achieve precise prediction result. After removing duplicated or invalid compounds, there were 1801 compounds in total. Of these, 669 CCLs with 402 compounds in CTRP dataset, 474 CCLs with 1,285 compounds in PRISM dataset and 786 CCLs with 320 compounds in GDSC dataset were used for subsequent drug prediction analysis (Fig. 3D). The ridge regression model located in the package *pRRophetic* was applied to perform the drug response prediction of clinical samples on the basis of their expression profiles, and the estimated AUC value of each compound among clinical samples was used as an evaluation indicator of drug sensitivity.

Before proceeding further, the results of drug response estimation have been validated computationally. Pazopanib, an oral small-molecule multi-kinase inhibitor for the treatment of advanced renal cell carcinoma, was used to evaluate whether the estimated drug sensitivity was consistent with its clinical efficacy. A retrospective cohort study found that the mutation status of *BAP1* has independent prognostic value in advanced RCC patients treated with first-line tyrosine kinase inhibitors [46]. Compared with WT patients, those patients harboring *BAP1* mutation performed reduced clinical benefit from pazopanib treatment, and exhibited worse PFS and overall survival (OS). Therefore, patients from the combined RNA-seq cohort were divided into two groups according to their alteration statuses of *BAP1* (altered versus unaltered: 84 versus 526). The Wilcoxon rank-sum test was applied to compare the estimated AUC values of pazopanib between two groups, and the result suggested that a significantly higher value of patients with mutant *BAP1* than WT ($P = 0.018$) (Fig. 3E), consistent with what pazopanib behaved clinically.

Constructing prediction model of *BAP1* mutation

On the basis of the combined RNA-seq cohort, the EN algorithm described in the Materials and Methods section was utilized to construct a robust model for predicting *BAP1* mutation status. The differential genes between the *BAP1* mutant and WT samples should be provided to this prediction model. Therefore, the *limma* package was applied to investigate the expression difference of these samples and differential genes were defined as *P adjust*< 0.05 and absolute log2 fold change (FC) > 1.

Survival analysis on 1,207 patients with both available prognosis and mutation data was conducted to investigate whether the functional status of *BAP1* was associated with the survival outcome of cancer

patients. The result denoted that there was a significant prognostic difference between the two groups, with longer median survival time (MST) in WT patients (MST=6.16 years, 95% confidence interval [CI]: 5.31-7.95 years) than in BAP1 mutant patients (MST=2.46 years, 95%CI: 2.00-3.52 years), which was consistent with the results of the MSKCC and the TCGA-KIRC cohorts (Additional file 1: Fig. S3).

To discern the characterization of biological processes affected by *BAP1* mutation, the enrichment analysis was performed using R package *GSVA*. The result showed that the up-regulated genes in *BAP1* mutant group were enriched in multiple carcinogenesis associated pathways, such as E2F targets, MTORC1 signaling and DNA repair, while the up-regulated genes in WT group were enriched in metabolism-associated pathways, such as pancreas beta cells, bile acid metabolism (Fig. 4A and Additional file 2: Table S5).

Based on *BAP1* mutation prediction model, the prediction accuracy achieved 93.1% in the training cohort (combined RNA-seq cohort) and 84.2% in the independent validation cohort (E-MTAB-1980) (Additional file 1: Fig. S4A and S4B). To evaluate the predictive ability of the prediction model, receiver operating characteristic (ROC) curve was applied using R package *pROC*, which a higher AUC indicates a preferable performance of the model. The AUC of this prediction model was 0.956 in training cohort and 0.895 in validation cohort (Additional file 1: Fig. S4C and S4D), suggesting that this model was efficient and robust enough for predicting *BAP1* alteration in other transcriptomic cohorts. Therefore, this model was used to identify estimated *BAP1* mutant samples from combined microarray cohort (E-MTAB-3267, E-MTAB-3218, E-MTAB-1980 and GSE29609).

Identification of potential therapeutic agents for *BAP1* mutant ccRCC

According to target annotation, 167 associated drugs were retained after mapping drugs to 69 unique targets in the DM-DG pairs. Differential drug response analyses between WT and mutant patients were conducted to further connect DMs with these DG-associated drugs. Compared with WT samples, only drugs with significantly lower estimated AUC values in mutated samples (logFC < 0 & *P* value<0.05) were considered as SL-associated drugs. There remain 149 DM-drug pairs and 49 DM-DG pairs meet the screening requirements, which were then visualized in a DM-DG-drug network (Fig. 4B and Additional file 2: Table S6). Among the final candidate SL pairs, the number of *BAP1* mutant gene pairs was far more than other DM-DG pairs, which provided more potential therapeutic agents for this kind of patients. Since *BAP1* mutated tumors were significantly associated with worse overall survival than tumors without mutated *BAP1* [6], it is essential to investigate specialized therapeutic agents for *BAP1* mutant ccRCC. Accordingly, the *BAP1* mutation was selected for further investigations regarding its therapeutic potential in renal cancer.

In the DM-DG-drug network, these analyses yielded 26 compounds with potential therapeutic effects for treating *BAP1*-mutant ccRCC. We compared the dependency scores of specific compound targets

between *BAP1* mutant and WT cells from RCC to validate the effect of these potential drugs (Fig. 4C-F). Although there was no statistically difference in results, CCLs with *BAP1* mutation still exhibited a trend toward lower dependency scores. Through integrating drug prediction results, survival and dependency analyses, it was found that *BRD4* and *PRKDC* could be the optimal targets for treating ccRCC patients with *BAP1* mutations (Fig. 5A). Nevertheless, above analyses alone cannot fully support the conclusion that the actual effect of compounds applied to tumors was consistent with the theoretical inference. Therefore, to explore the potential effect of these compounds in treating ccRCC, the multiple perspective approaches for drug prediction were adopted. First, the CMap analysis was utilized to find medicines whose drug signatures, namely drug-induced profiles of expression changes, were opposite to the *BAP1* mutant expression pattern. A total of three compounds, including ZSTK-474, BI-2536 and PI-103, had CMap scores less than -80, representing that these drugs might have therapeutic effects in patients with *BAP1* mutations. Second, we calculated the expression differences of candidate DG between normal and tumor tissue, and compounds with higher fold change values were considered as greater potential agents for ccRCC treatment. Thirdly, through searching relevant literature about compounds in Pubmed (https://pubmed.ncbi.nlm.nih.gov/), we found out the experimental and clinical evidence of candidates in treating ccRCC. Lastly, the dependency analysis of DGs across kidney CCLs was also conducted, and lower CERES or DEMETER scores denote that relevant genes are more likely to be essential for CCLs survival. All results were presented in Fig. 5B and Additional file 2: Table S7. In general, BI-2536 and PI-103, which had robust abilities in vitro and in silico evidence, were considered as the best therapeutic compounds for *BAP1* specific ccRCC treatment.

In addition, the independent dataset, which comprised molecular profiles and mutation data of 246 ccRCC patients from the combined microarray cohort, was also used for further external validation. Through comparing the estimated AUC values of two specific agents (BI-2536 and PI-103) between *BAP1* mutant and WT groups, the result suggested that mutant group was indeed more sensitive to both BI-2536 and PI-103 than WT group, highly consistent with the results of our in silico prediction (Fig. 5C, 5D and Additional file 2: Table S7).

Discussion

A recently accepted concept of tumorigenesis and progression is that tumor cells are susceptible to mutation events, thus they depend on other genes to gain survival advantages. Considering a pivotal challenge to rescue the activity of driver targets, it is urgent to discover alternative approaches. Fortunately, pharmaceutical agents based on SL strategy provide novel insight for precisely killing tumor cells with certain mutations. The PARP inhibitor Olaparib is the first drug to be clinically used in treating breast cancer patients with *BRCA1/2* mutation based on the SL interaction mechanism [47]. Although pan-cancer analysis has obtained considerable results [10, 48], the practical application value in ccRCC patients may be limited due to their distinct metabolism process, proliferative characteristic and genetic feature.

The applications of RNA interference (RNAi) and clustered regularly interspaced short palindromic repeats (CRISPR) are preferable choices to identify SL pairs, but such methods are expensive and only suitable for screening partners of few fascinating driver genes [8, 49]. Currently, given the easily accessible genomic data, using computational procedure is attractive to predict SL pairs. In the current study, we performed SL interaction analysis in the most comprehensive metadata set of ccRCC so far, which included 610 patients from five clinical cohorts with available expression and mutation data, to predict the potential gene pairs. The first predictive method, differential gene expression, assumes that most mutations of driver genes result in loss-of-function and hence allows the tumor cells to compensatively up-regulate the expression of the SL partners [9]. The second predictive method, pairwise gene co-expression, depends on the concept that SL pairs seem to exert related biological functions and co-express in WT tumor samples [9]. The third predictive method, functional similarity, indicates that gene pairs with SL interaction are likely to engage in similar biological process, thus their locations in GO topological network should be neighboring. The last one, mutual exclusivity, is based on the notion that inhibition of two genes with SL interplay can reduce tumor cells vitality and hence two genes of tumor samples express in a mutually exclusive manner [50].

Classifying the genomic characteristics provides a brilliant prospect for the occurrence, progression and precise treatment of RCC. That is, *VHL* mutation acted as an initiative event to induce tumor occurrence, while *PBRM1*, *BAP1* and *SETD2* cause DNA repair defect and cell overgrowth. Subsequently, the effective pathways, such as PI3K-mTOR activation, confer tumor cells the potential to evade death signals and metastasis [6]. In this study, chromatin remodeling gene *BAP1* accounts for 59.7% of potential SL-based driver genes, followed by another frequent mutating gene *PBRM1* (23.6%). It is revealed that *BAP1* and *PBRM1*, residing closely on chromosome 3p, are frequently mutated (approximately 10% and 40%, respectively) in RCC patients [51–53]. Several studies have proved the crucial role of *BAP1* and *PBRM1* in tumor development. Briefly, BAP1 interacts with BRCA1/ BARD1 complex to regulate crucial biological processes, such as chromatin modification, DNA damage repair and cell cycle control [54, 55]. Depletion of *BAP1* was associated with aggressive histological grade [52], advanced tumor stage [56] and poor prognosis [54]. Additionally, *BAP1* mutation was correlated with high genome instability index (GII) and low intratumoural heterogeneity (ITH), conferring the adaptive advantage and single lethal target to ccRCC clone [57]. In regards to *PBRM1*, its depletion promoted the upregulation of *HIF-1α*, *STAT3* and the activation of mTOR signaling induced by *VHL* mutation [58]. Such phenomenon may explain that patients with *BAP1* mutation experienced a worse outcome than patients with *PBRM1* mutation after receiving first-line VEGFR inhibitor everolimus and mTOR inhibitor sunitinib treatment [53].

To explore available compounds for clinical usefulness, we further estimated drug response of clinical samples from pharmacogenomics profile databases CTRP, PRISM and GDSC. The estimated drug sensitivity of Bromodomain Containing 4 (BRD4) inhibitor BI-2536, phosphoinosmde-3-kinase (PI3K)/mammalian target of rapamycin (mTOR) inhibitor PI-103, and PI3K specific inhibitor ZSTK474 in *BAP1* mutated samples are attractive for further study due to their desirable matching scores. In this study, BI-2536 showed a high drug sensitivity against *BAP1* mutated samples by inhibiting BRD4 function. Among ccRCC patients with *BAP1* mutation, the up-regulated expression of *BRD4* was

associated with poor prognosis, indicating a possible benefit of BRD4 inhibition in *BAP1* mutated samples. It is well-known that BRD4, an important component of the bromodomain and extra terminal (BET) protein family, shares similar functions with BAP1 in chromatin remodeling and transcriptional regulation [59, 60]. The up-regulation of *BRD4* expression was found in RCC tissues, and associated with advanced histological stage and lymph node metastasis, while knockdown of *BRD4* reduced cell vitality and inhibited tumor growth [61]. The BRD4 inhibitor JQ-1 enhanced the anti-tumor activity of the mTOR inhibitor Palomid 529 in RCC cells [62]. Malignant peripheral nerve sheath tumor with *PRC2* loss-of-function was sensitive to BRD4 inhibitor, suggesting a promising therapeutic approach of SL-based BRD4 inhibition [63]. The dual PI3K/mTOR inhibitor PI-103 is available to treat various tumor types. For example, the inhibitory ability of SCD-1 interference on cell proliferation and migration of RCC cells was amplified by PI-103 [64]. Combination of PI-103 and mTOR inhibitor rapamycin performed a better therapeutic effect than single agents in human ovarian and prostate cancer cells, and can effectively prevent rebound activation of the Akt pathway treated after rapamycin treatment [65]. In addition to the PI-103, ZSTK474, another inhibitor that specifically targets PI3K, also received a high score in our analysis. In vitro experiments have shown that it can inhibit the proliferation of tumor cells through interfering cell G0/1 stage arrest [66, 67]. It is exciting that ZSTK474 induced the degradation of multidrug efflux pumps ABCB1 and ABCG2 so as not to be affected by the efflux effect of resistant cancer cells [68]. Furthermore, ZSTK474 exhibited antiangiogenic activity via downregulating HIF-1α and VEGF, and suppressed renal cancer growth in a xenograft model [69]. Generally, above evidence of these three compounds indirectly proved the reasonability of our computational pipeline and the reliability of the prediction results.

This study still has several limitations. First, several studies employed pairwise survival analysis to SL identification [9, 70], which was not included in our screening criteria, for the reason that the relatively low mutation frequency of crucial driver genes like *BAP1* and some inaccessible survival data of cohorts would reduce the statistical power and thus ignore several important SL interactions. Second, despite robust evidence from pharmaceutical database, there is still a lack of experimental validation. Related experiments are needed in the future to support our conclusions. Third, BI-2536 is also considered as PLK1 inhibitor [71], so further exploration of the target of BI-2536 is essential to elucidate its anti-cancer mechanism in ccRCC.

Conclusion

In conclusion, capitalizing on extensive screening data combined with molecular and clinical data from multiple cohorts, this study developed a novel computational-based strategy to identify SL pairs for ccRCC patients harboring genetically mutation as well as some potential therapeutic agents for *BAP1* mutated patients. The potential SL-associated partners for *BAP1* and *PBRM1*, two frequent altered genes, have complemented the current *VHL*-predominant research and mapped a comprehensive landscape for SL interaction in ccRCC, which might help to deepen our understanding of ccRCC mutation patterns and provide an alternative strategy of personalized renal cancer treatment.

Abbreviations

ccRCC: Clear cell renal cell carcinoma; SL: Synthetic lethality; RCC: Renal cell carcinoma; DM: Driver mutation; PARP: Poly (ADP-ribose) polymerase; VHL: Von Hippel-Lindau; BAP1: BRCA1 associated protein 1; HIF: Hypoxia inducible factor; TCGA-KIRC: The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma; RECA-EU: Renal cell cancer-EU; CM-009: CheckMate 009; CM-010: CheckMate 010; CM-025: CheckMate 025; DG: Druggable gene; GO: Gene Ontology; FSS: Functional similarity score; MF: Molecular function; CC: Cellular component; NBS: Network-based stratification; PFS: Progression-free survival; CCLs: Cancer cell lines; CTRP: Cancer Therapeutics Response Portal; GDSC: Genomics of Drug Sensitivity in Cancer; AUC: Area under curve; WT: Wild-type; OS: overall survival; EN: Elastic net; MST: Median survival time; CMap: Connectivity map; RNAi: RNA interference; CRISPR: clustered regularly interspaced short palindromic repeats; GII: Genome instability index; ITH: Intratumoural heterogeneity; BRD4: Bromodomain Containing 4; PI3K: phosphoinosmde-3-kinase; mTOR: mammalian target of rapamycin; BET: bromodomain and extra terminal; ICGC: International Cancer Genome Consortium; TPM: transcripts per million; SNVs: single nucleotide variants; INDELs: small insertions/deletions; CNVs: copy number variants; RMA: robust multiarray average; GEO: Gene Expression Omnibus; CCLE: Cancer Cell Line Encyclopedia; KNN: K-nearest neighbors; DepMap: Dependency Map; ROC: receiver operating characteristic; NTP: nearest template prediction; SS: semantic similarity; RAS: rank aggregation score; GSVA: gene set variation analysis.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

Z.C.L., D.X.L and L.M.Z were responsible for the conception and design. Z.C.L. and D.X.L contributed to the acquisition of the data. Z.C.L., D.X.L performed the execution, analysis and interpretation of the data. Z.C.L., D.X.L and L.M.Z visualized and managed the results. Z.C.L. and D.X.L contributed to the drafting of the manuscript. C.Y., B.G., Y.L., D.Y.C., C.W., K.W., Z.Y.X., Z.C. and C.L. were responsible for the supervision and manuscript revising. All authors read and approved the final manuscript.

Funding

This work is supported by grants from the National Natural Science Foundation of China (81873625 to Z.C., 81802931 to C.L.).

Availability of data and materials

The datasets supporting the conclusions of this article are included within the additional file of this article. If any other data are needed, please contact the corresponding author.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors give consent for publication.

Competing interests

The authors declare that they have no conflict of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71:209-249. https://doi.org/10.3322/caac.21660.
2. Stewart GD, Omahony FC, Powles T, Riddick ACP, Harrison DJ, Faratian D. What can molecular pathology contribute to the management of renal cell carcinoma. Nat Rev Urol. 2011;8:255-265. https://doi.org/10.1038/nrurol.2011.43.
3. Lam JS, Leppert JT, Figlin RA, Belldegrun AS. Surveillance following radical or partial nephrectomy for renal cell carcinoma. Curr Urol Rep. 2005;6:7-18. https://doi.org/10.1007/s11934-005-0062-x.
4. Linehan WM. Genetic basis of kidney cancer: role of genomics for the development of disease-based therapeutics. Genome Res. 2012;22:2089-2100. https://doi.org/10.1101/gr.131110.111.
5. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. BMJ. 2014;349:g4797. https://doi.org/10.1136/bmj.g4797.
6. Dizman N, Philip EJ, Pal SK. Genomic profiling in renal cell carcinoma. Nat Rev Nephrol. 2020;16:435-451. https://doi.org/10.1038/s41581-020-0301-x.
7. Thompson JM, Nguyen QH, Singh M, Pavesic MW, Nesterenko I, Nelson LJ, et al. Rho-associated kinase 1 inhibition is synthetically lethal with von Hippel-Lindau deficiency in clear cell renal cell carcinoma. Oncogene. 2017;36:1080-1089. https://doi.org/10.1038/onc.2016.272.
8. Sun N, Petiwala S, Lu C, Hutti JE, Hu M, Hu M, et al. VHL Synthetic Lethality Signatures Uncovered by Genotype-Specific CRISPR-Cas9 Screens. CRISPR J. 2019;2:230-245. https://doi.org/10.1089/crispr.2019.0018.
9. Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell. 2014;158:1199-1209. https://doi.org/10.1016/j.cell.2014.07.027.
10. Sinha S, Thomas D, Chan S, Gao Y, Brunen D, Torabi D, et al. Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. Nat Commun. 2017;8:15580. https://doi.org/10.1038/ncomms15580.

11. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499:43-49. https://doi.org/10.1038/nature12222.
12. Choueiri TK, Fishman MN, Escudier B, McDermott DF, Drake CG, Kluger H, et al. Immunomodulatory Activity of Nivolumab in Metastatic Renal Cell Carcinoma. Clin Cancer Res. 2016;22:5461-5471. https://doi.org/10.1158/1078-0432.CCR-15-2839.
13. Motzer RJ, Rini BI, Mcdermott DF, Redman BG, Kuzel TM, Harrison MR, et al. Nivolumab for Metastatic Renal Cell Carcinoma: Results of a Randomized Phase II Trial. J Clin Oncol. 2014;33:1430-1437.
14. Motzer RJ, Tannir NM, McDermott DF, Aren Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. N Engl J Med. 2018;378:1277-1290. https://doi.org/10.1056/NEJMoa1712126.
15. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med. 2020;26:909-918. https://doi.org/10.1038/s41591-020-0839-y.
16. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882-883. https://doi.org/10.1093/bioinformatics/bts034.
17. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20:307-315. https://doi.org/10.1093/bioinformatics/btg405.
18. Di Lena P, Sala C, Prodi A, Nardini C. Missing value estimation methods for DNA methylation data. Bioinformatics. 2019;35:3786-3793. https://doi.org/10.1093/bioinformatics/btz134.
19. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat Genet. 2017;49:1779-1784. https://doi.org/10.1038/ng.3984.
20. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a Cancer Dependency Map. Cell. 2017;170:564-576.e516. https://doi.org/10.1016/j.cell.2017.06.010.
21. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33:1-22.
22. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47. https://doi.org/10.1093/nar/gkv007.
23. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77. https://doi.org/10.1186/1471-2105-12-77.
24. Bashashati A, Haffari G, Ding J, Ha G, Lui K, Rosner J, et al. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. Genome Biol. 2012;13:R124. https://doi.org/10.1186/gb-2012-13-12-r124.

25. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. Genome Biol. 2010;11:R53. https://doi.org/10.1186/gb-2010-11-5-r53.
26. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The Drug Repurposing Hub: a next-generation drug library and information resource. Nat Med. 2017;23:405-408. https://doi.org/10.1038/nm.4306.
27. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46:D1074-D1082. https://doi.org/10.1093/nar/gkx1037.
28. Canisius S, Martens JW, Wessels LF. A novel independence test for somatic alterations in cancer shows that biology drives mutual exclusivity but chance explains most co-occurrence. Genome Biol. 2016;17:261. https://doi.org/10.1186/s13059-016-1114-x.
29. Huang JK, Jia T, Carlin DE, Ideker T. pyNBS: a Python implementation for network-based stratification of tumor mutations. Bioinformatics. 2018;34:2859-2861. https://doi.org/10.1093/bioinformatics/bty186.
30. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nat Methods. 2013;10:1108-1115. https://doi.org/10.1038/nmeth.2651.
31. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11:367. https://doi.org/10.1186/1471-2105-11-367.
32. Hoshida Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. PLoS One. 2010;5:e15543. https://doi.org/10.1371/journal.pone.0015543.
33. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010;26:976-978. https://doi.org/10.1093/bioinformatics/btq064.
34. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012;28:573-580. https://doi.org/10.1093/bioinformatics/btr709.
35. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science. 2003;302:249-255. https://doi.org/10.1126/science.1087447.
36. Azuaje F. Computational models for predicting drug responses in cancer research. Brief Bioinform. 2017;18:820-829. https://doi.org/10.1093/bib/bbw065.
37. Geeleher P, Zhang Z, Wang F, Gruener RF, Nath A, Morrison G, et al. Discovering novel pharmacogenomic biomarkers by imputing drug response in cancer patients from large genomics studies. Genome Res. 2017;27:1743-1751. https://doi.org/10.1101/gr.221077.117.
38. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417-425. https://doi.org/10.1016/j.cels.2015.12.004.
39. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. https://doi.org/10.1186/1471-2105-14-7.

40. Lee JS, Nair NU, Dinstag G, Chapman L, Chung Y, Wang K, et al. Synthetic lethality-mediated precision oncology via the tumor transcriptome. Cell. 2021;184:2487-2502.e2413. https://doi.org/10.1016/j.cell.2021.03.030.
41. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011;469:539-542. https://doi.org/10.1038/nature09639.
42. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010;463:360-363. https://doi.org/10.1038/nature08672.
43. Rini B, Goddard A, Knezevic D, Maddala T, Zhou M, Aydin H, et al. A 16-gene assay to predict recurrence after surgery in localised renal cell carcinoma: development and validation studies. Lancet Oncol. 2015;16:676-685. https://doi.org/10.1016/s1470-2045(15)70167-1.
44. Brooks SA, Brannon AR, Parker JS, Fisher JC, Sen O, Kattan MW, et al. ClearCode34: A prognostic risk predictor for localized clear cell renal cell carcinoma. Eur Urol. 2014;66:77-84. https://doi.org/10.1016/j.eururo.2014.02.035.
45. Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. J Clin Oncol. 1999;17:2530-2540. https://doi.org/10.1200/jco.1999.17.8.2530.
46. Voss MH, Reising A, Cheng Y, Patel P, Marker M, Kuo F, et al. Genomically annotated risk model for advanced renal-cell carcinoma: a retrospective cohort study. Lancet Oncol. 2018;19:1688-1698. https://doi.org/10.1016/s1470-2045(18)30648-x.
47. Mateo J, Lord CJ, Serra V, Tutt A, Balmaña J, Castroviejo-Bermejo M, et al. A decade of clinical development of PARP inhibitors in perspective. Ann Oncol. 2019;30:1437-1447. https://doi.org/10.1093/annonc/mdz192.
48. Behan FM, Iorio F, Picco G, Goncalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature. 2019;568:511-516. https://doi.org/10.1038/s41586-019-1103-9.
49. Nicholson HE, Tariq Z, Housden BE, Jennings RB, Stransky LA, Perrimon N, et al. HIF-independent synthetic lethality between CDK4/6 inhibition and VHL loss across species. Sci Signal. 2019;12. https://doi.org/10.1126/scisignal.aay0482.
50. Srihari S, Singla J, Wong L, Ragan MA. Inferring synthetic lethal interactions from mutual exclusivity of genetic events in cancer. Biol Direct. 2015;10:57. https://doi.org/10.1186/s13062-015-0086-1.
51. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Rep. 2018;23:313-326 e315. https://doi.org/10.1016/j.celrep.2018.03.075.
52. Guo G, Gui Y, Gao S, Tang A, Hu X, Huang Y, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. Nat Genet. 2011;44:17-19. https://doi.org/10.1038/ng.1014.

53. Hsieh JJ, Chen D, Wang PI, Marker M, Redzematovic A, Chen YB, et al. Genomic Biomarkers of a Randomized Trial Comparing First-line Everolimus and Sunitinib in Patients with Metastatic Renal Cell Carcinoma. Eur Urol. 2017;71:405-414. https://doi.org/10.1016/j.eururo.2016.10.007.
54. Kapur P, Peña-Llopis S, Christie A, Zhrebker L, Pavía-Jiménez A, Rathmell WK, et al. Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. Lancet Oncol. 2013;14:159-167. https://doi.org/10.1016/s1470-2045(12)70584-3.
55. Nishikawa H, Wu W, Koike A, Kojima R, Gomi H, Fukuda M, et al. BRCA1-associated protein 1 interferes with BRCA1/BARD1 RING heterodimer activity. Cancer Res. 2009;69:111-119. https://doi.org/10.1158/0008-5472.Can-08-3355.
56. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333-339. https://doi.org/10.1038/nature12634.
57. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. Cell. 2018;173:595-610 e511. https://doi.org/10.1016/j.cell.2018.03.043.
58. Nargund AM, Pham CG, Dong Y, Wang PI, Osmangeyoglu HU, Xie Y, et al. The SWI/SNF Protein PBRM1 Restrains VHL-Loss-Driven Clear Cell Renal Cell Carcinoma. Cell Rep. 2017;18:2893-2906. https://doi.org/10.1016/j.celrep.2017.02.074.
59. Floyd SR, Pacold ME, Huang Q, Clarke SM, Lam FC, Cannell IG, et al. The bromodomain protein Brd4 insulates chromatin from DNA damage signalling. Nature. 2013;498:246-250. https://doi.org/10.1038/nature12147.
60. Schroder S, Cho S, Zeng L, Zhang Q, Kaehlcke K, Mak L, et al. Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. J Biol Chem. 2012;287:1090-1099. https://doi.org/10.1074/jbc.M111.282855.
61. Wu X, Liu D, Gao X, Xie F, Tao D, Xiao X, et al. Inhibition of BRD4 Suppresses Cell Proliferation and Induces Apoptosis in Renal Cell Carcinoma. Cell Physiol Biochem. 2017;41:1947-1956. https://doi.org/10.1159/000472407.
62. Xing ZY, Wang Y, Cheng L, Chen J, He XZ, Xing W. Bromodomain-Containing Protein 4 (BRD4) Inhibition Sensitizes Palomid 529-Induced Anti-Renal Cell Carcinoma Cell Activity in Vitro and in Vivo. Cell Physiol Biochem. 2018;50:640-653. https://doi.org/10.1159/000494185.
63. De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, et al. PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. Nature. 2014;514:247-251. https://doi.org/10.1038/nature13561.
64. Wang H, Zhang Y, Lu Y, Song J, Huang M, Zhang J, et al. The role of stearoyl-coenzyme A desaturase 1 in clear cell renal cell carcinoma. Tumour Biol. 2016;37:479-489. https://doi.org/10.1007/s13277-015-3451-x.

65. Mazzoletti M, Bortolin F, Brunelli L, Pastorelli R, Di Giandomenico S, Erba E, et al. Combination of PI3K/mTOR inhibitors: antitumor activity and molecular correlates. Cancer Res. 2011;71:4573-4584. https://doi.org/10.1158/0008-5472.CAN-10-4322.
66. Dan S, Yoshimi H, Okamura M, Mukai Y, Yamori T. Inhibition of PI3K by ZSTK474 suppressed tumor growth not via apoptosis but G0/G1 arrest. Biochem Biophys Res Commun. 2009;379:104-109. https://doi.org/10.1016/j.bbrc.2008.12.015.
67. Dan S, Okamura M, Mukai Y, Yoshimi H, Inoue Y, Hanyu A, et al. ZSTK474, a specific phosphatidylinositol 3-kinase inhibitor, induces G1 arrest of the cell cycle in vivo. Eur J Cancer. 2012;48:936-943. https://doi.org/10.1016/j.ejca.2011.10.006.
68. Muthiah D, Callaghan R. Dual effects of the PI3K inhibitor ZSTK474 on multidrug efflux pumps in resistant cancer cells. Eur J Pharmacol. 2017;815:127-137. https://doi.org/10.1016/j.ejphar.2017.09.001.
69. Kong D, Okamura M, Yoshimi H, Yamori T. Antiangiogenic effect of ZSTK474, a novel phosphatidylinositol 3-kinase inhibitor. Eur J Cancer. 2009;45:857-865. https://doi.org/10.1016/j.ejca.2008.12.007.
70. Lee JS, Das A, Jerby-Arnon L, Arafeh R, Auslander N, Davidson M, et al. Harnessing synthetic lethality to predict the response to cancer treatment. Nat Commun. 2018;9:2546. https://doi.org/10.1038/s41467-018-04647-1.
71. Ding Y, Huang D, Zhang Z, Smith J, Petillo D, Looyenga BD, et al. Combined gene expression profiling and RNAi screening in clear cell renal cell carcinoma identify PLK1 and other therapeutic kinase targets. Cancer Res. 2011;71:5225-5234. https://doi.org/10.1158/0008-5472.CAN-11-0076.

Figures

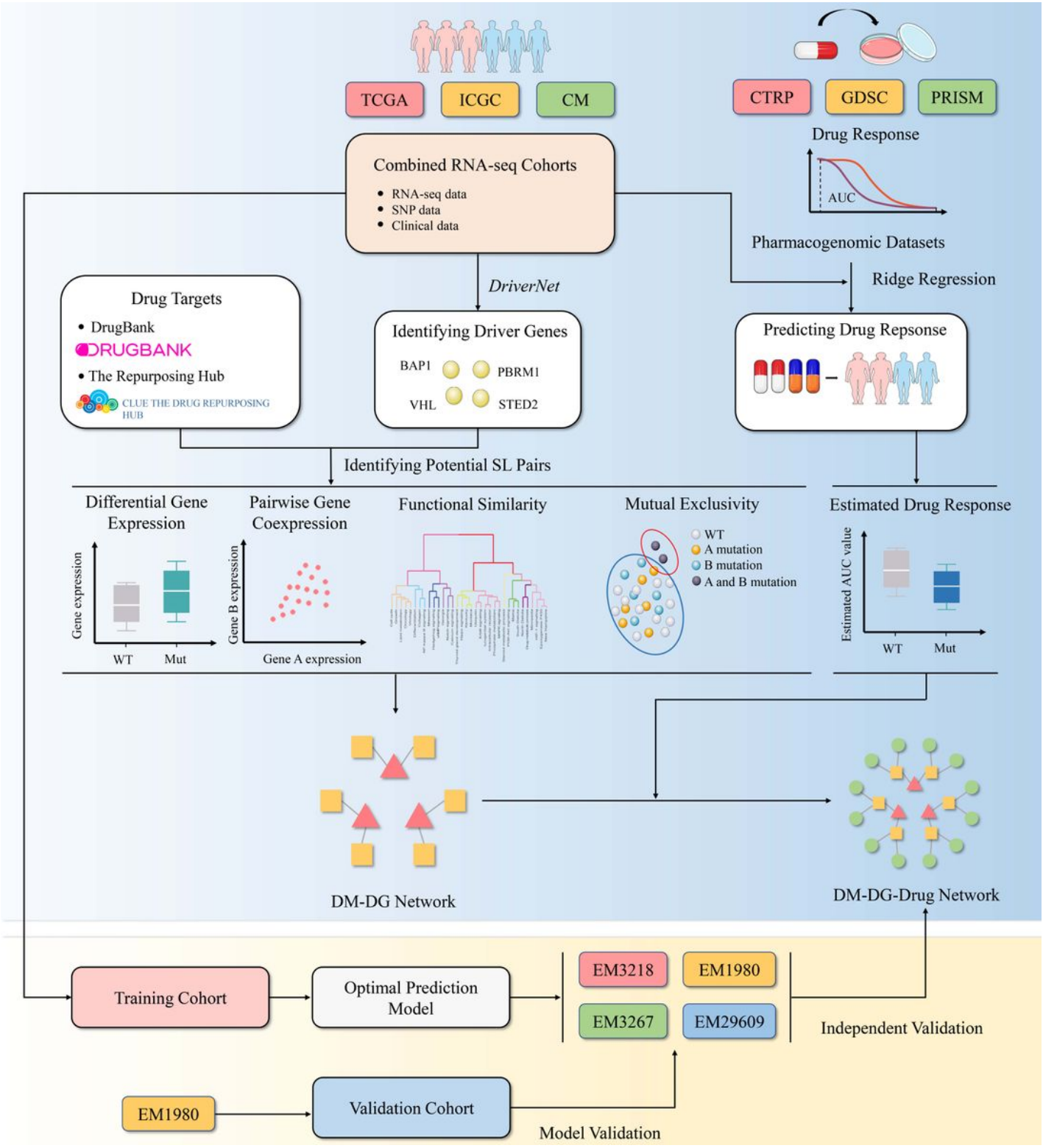

Figure 1

Flow chart of identification of potential synthetic lethal interactions and construction of DM-DG-drug networks.

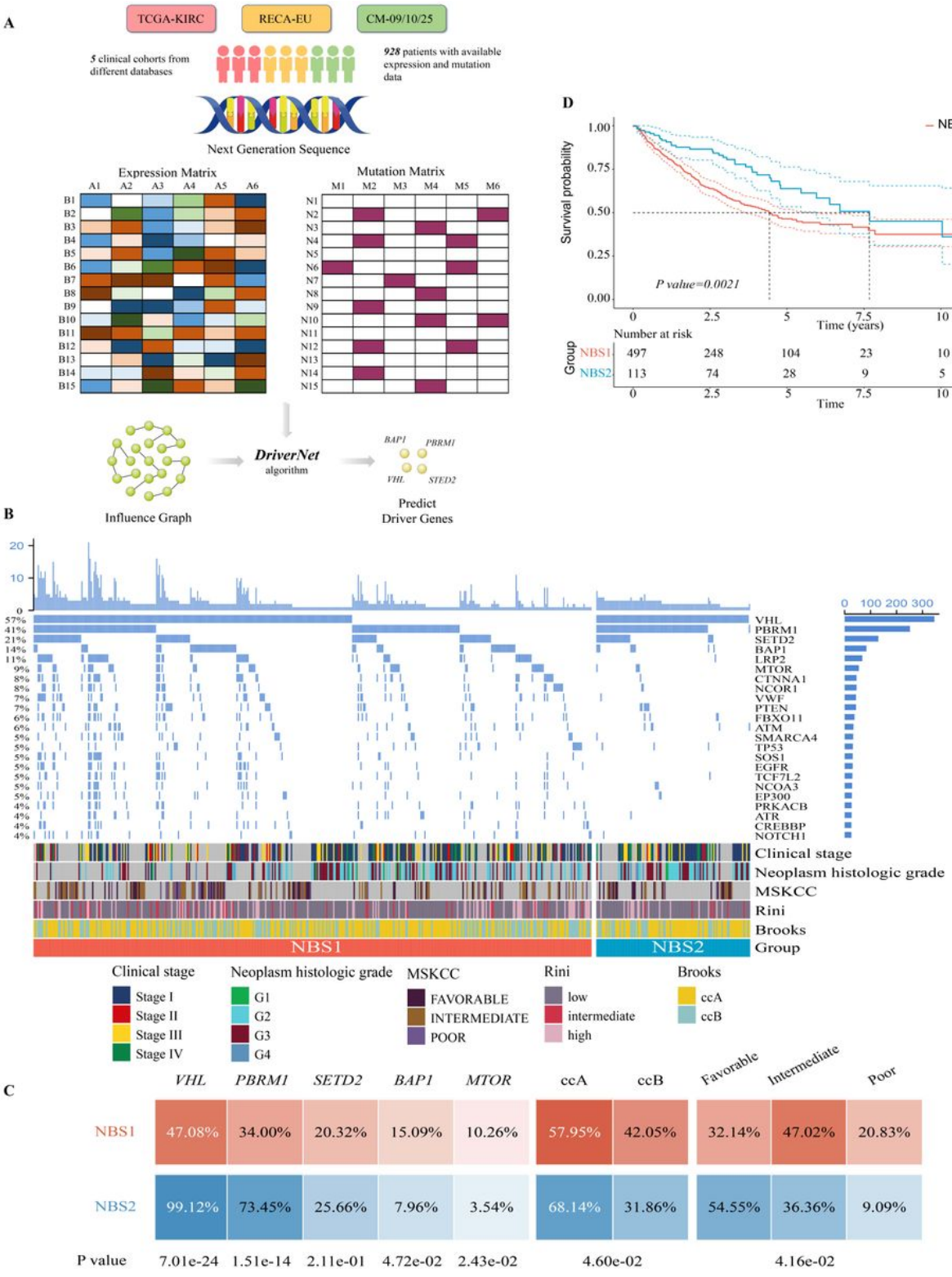

Figure 2

Identifying driver genes and subclass characteristics in clear cell renal cell carcinoma. (A). Overview of driver genes identification via DriverNet analysis in clinical cohorts. (B). The mutation profiles of subclasses classified by network-based stratification (NBS). Characteristics of clinical stage, histological grade, previously reported transcriptome-based molecular subclasses (MSKCC, Rini and Brooks) between two subclasses were presented simultaneously. (C). Difference in mutation frequency of driver genes,

molecular characteristics stratified by Brooks and MSKCC between two subclasses. Fisher's exact tests were applied to compared the statistical differences. (D). Kaplan–Meier survival curve of two subclasses. Statistical difference was calculated by log-rank test.

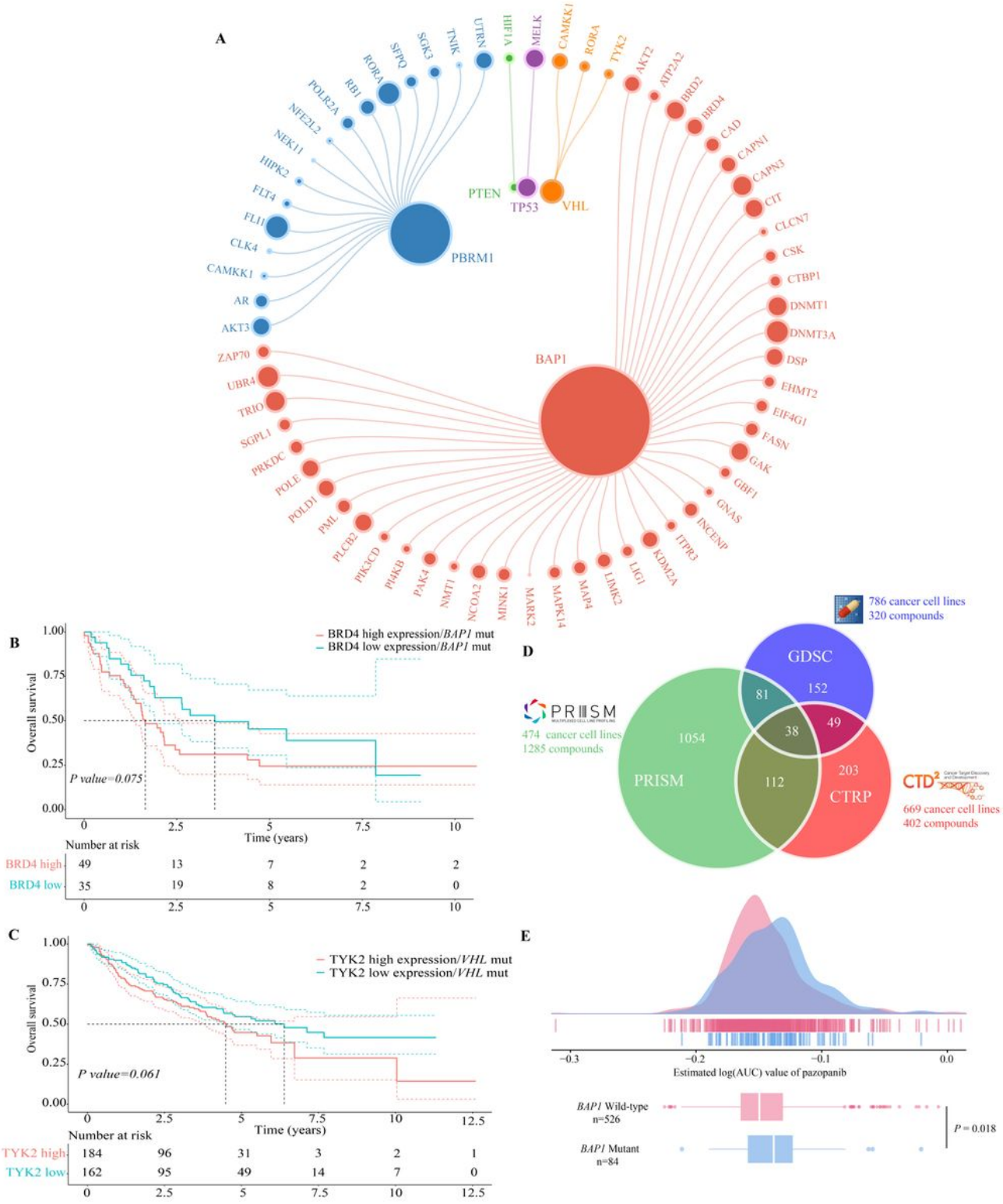

Figure 3

Exploring feasibility of druggable genes in treating driver mutation-specific clear cell renal cell carcinoma patients. (A). The bipartite network of representative DM-DG interactions. (B). Overall survival of distinct

BRD4 expression profiles in BAP1 mutated patients. (C). Overall survival of distinct TYK2 expression profiles in VHL mutated patients. (D). The venn graph for summarizing the available cancer cell lines and compounds in CTRP, PRISM and GDSC pharmacogenomic datasets. (E). Comparing estimated drug sensitivity (LogAUC) of pazopanib between BAP1 mutated and wild-type samples.

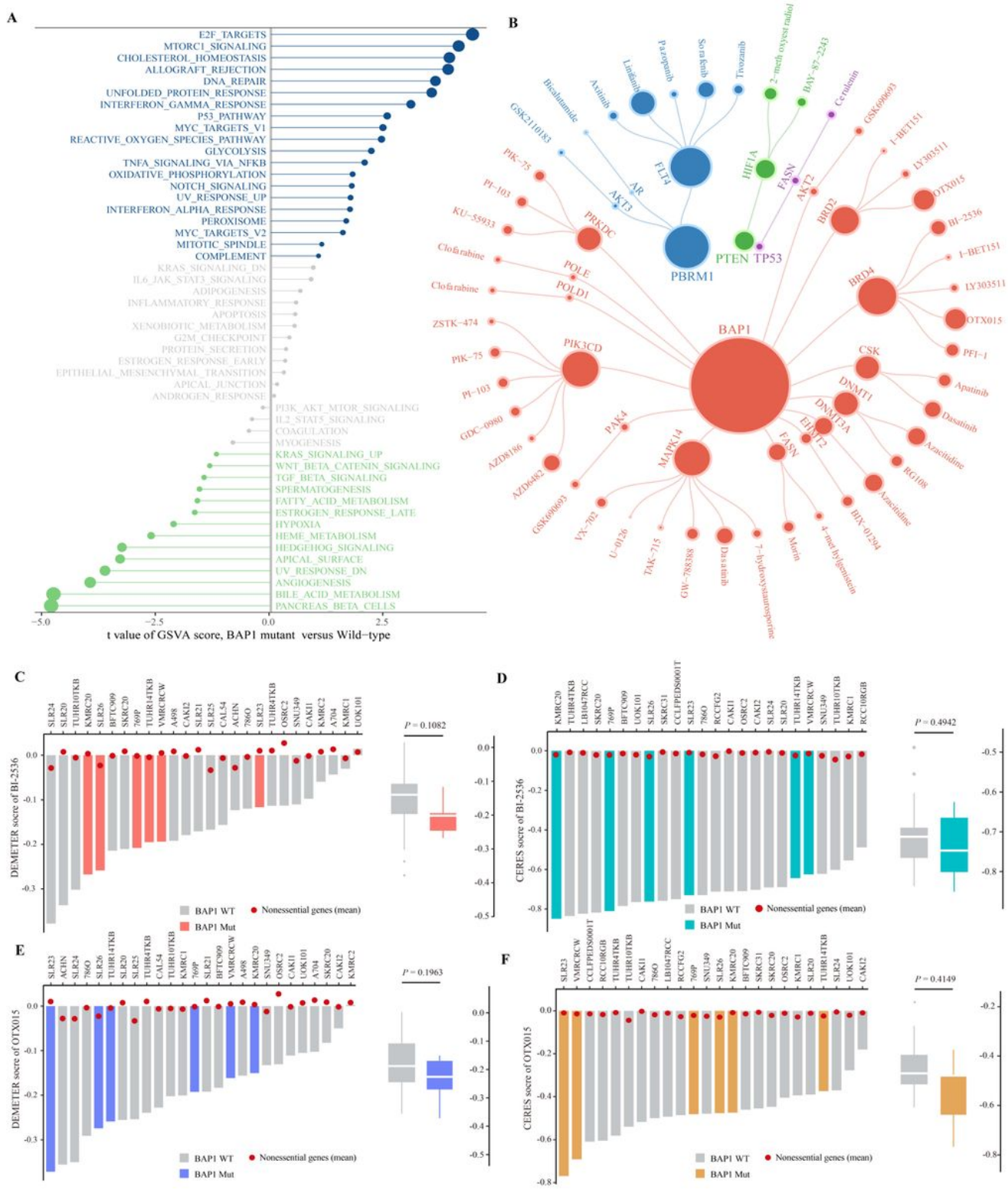

Figure 4

Determining sensitivities of identified drugs on renal cancer cell lines. (A). Gene set enrichment analysis between BAP1 mutated and wild-type groups. Blue dots indicate BAP1 mutant-enriched pathway, while red dots indicate wild type-associated pathways. (B). The bipartite network of representative TSG-DT-drug interactions. (C). The DEMETER scores derived from RNAi screens of BI-2536 across 24 liver CCLs. (D). The CERES scores derived from CRISPR knockout screens of BI-2536 across 26 liver CCLs. (E). The DEMETER scores derived from RNAi screens of OTX015 across 24 liver CCLs. (F). The CERES scores derived from CRISPR knockout screens of OTX015 across 26 liver CCLs.

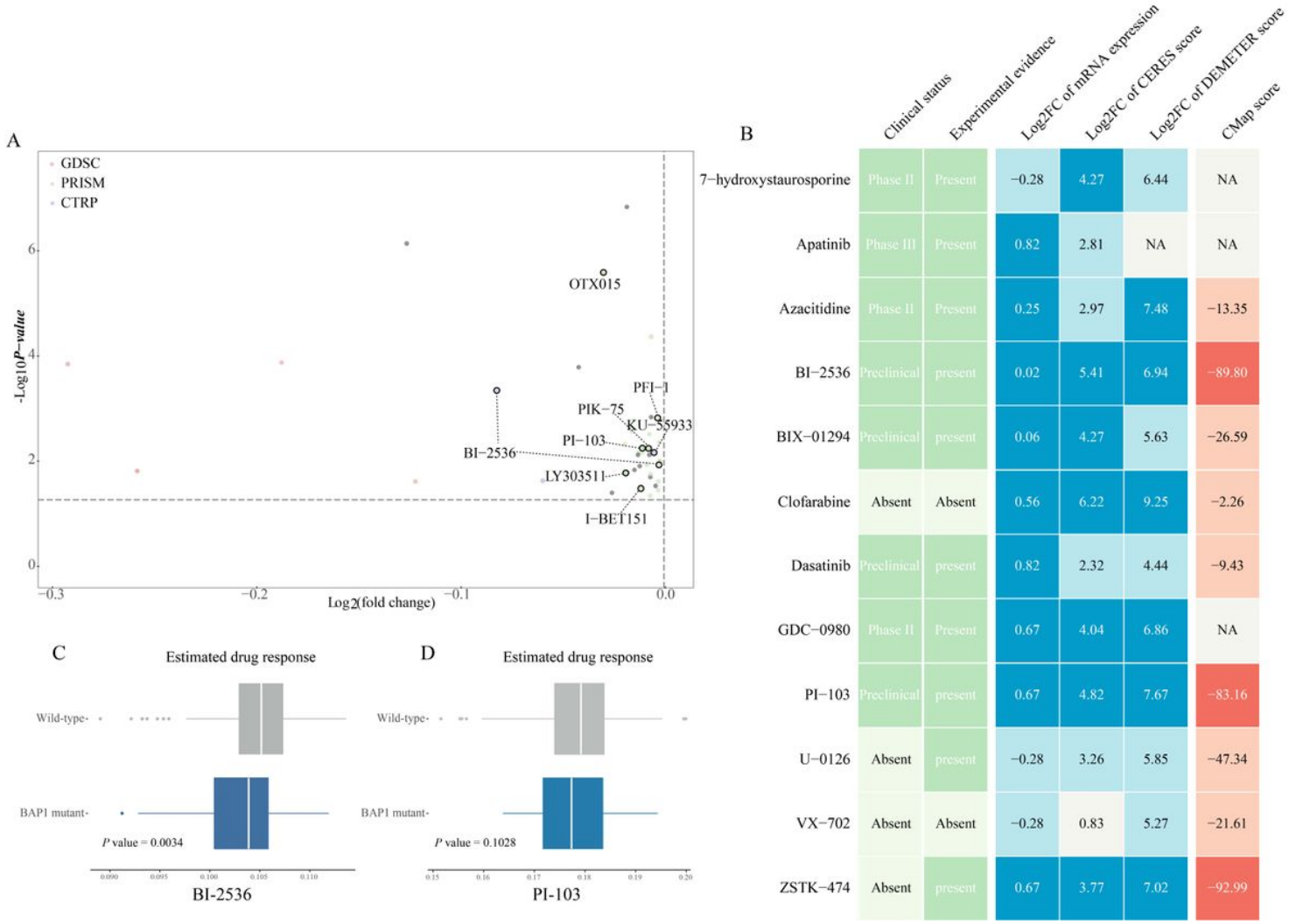

Figure 5

Estimating drug responses of BI-2536 and PI-103 across BAP1 mutated renal cancer patients. (A). Differential drug response analyses of identified 26 compounds with potential therapeutic efficacies on BAP1-mutant ccRCC. The BRD4 and PRKDC inhibitors with significant response differences between BAP1 mutant and wild type groups were labeled on the plot. (B). Summarizing the current evidences, target gene expression, drug dependency and CMap analysis of candidate drugs. (C). Estimating the drug responses of BI-2536 and PI-103 in treating BAP1 mutated and wild-type RCC patients.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf
- Additionalfile2.xlsx