Prediction of drug candidates for clear cell renal cell carcinoma using a systems biology-based drug repositioning approach

Prediction of drug candidates for clear cell renal cell carcinoma using a systems biology-based drug repositioning approach

Xiangyu Li,[a,b] Koeun Shong,[b] Woonghee Kim,[b] Meng Yuan,[b] Hong Yang,[b] Yusuke Sato,[c,d] Haruki Kume,[d] Seishi Ogawa,[c,e] Hasan Turkez,[f] Saeed Shoaie,[b,g] Jan Boren,[h] Jens Nielsen,[i,j] Mathias Uhlen,[b] Cheng Zhang,[b,k]* and Adil Mardinoglu [b,g]*

[a]Bash Biotech Inc, 600 est Broadway, Suite 700, San Diego, CA 92101, USA
[b]Science for Life Laboratory, KTH-Royal Institute of Technology, Stockholm SE-17165, Sweden
[c]Department of Pathology and Tumor Biology, Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto 606-8501, Japan
[d]Department of Urology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8654, Japan
[e]Centre for Hematology and Regenerative Medicine, Department of Medicine, Karolinska Institute, Stockholm SE-17177, Sweden
[f]Department of Medical Biology, Faculty of Medicine, Atatürk University, Erzurum 25240, Turkey
[g]Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London SE1 9RT, UK
[h]Department of Molecular and Clinical Medicine, University of Gothenburg, Sahlgrenska University Hospital, Gothenburg SE-41345, Sweden
[i]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg SE-41296, Sweden
[j]BioInnovation Institute, Copenhagen N DK-2200, Denmark
[k]Key Laboratory of Advanced Drug Preparation Technologies, School of Pharmaceutical Sciences, Ministry of Education, Zhengzhou University, Zhengzhou 450001, China

eBioMedicine 2022;78: 103963
Published online xxx
https://doi.org/10.1016/j.ebiom.2022.103963

Summary

Background The response rates of the clinical chemotherapies are still low in clear cell renal cell carcinoma (ccRCC). Computational drug repositioning is a promising strategy to discover new uses for existing drugs to treat patients who cannot get benefits from clinical drugs.

Methods We proposed a systematic approach which included the target prediction based on the co-expression network analysis of transcriptomics profiles of ccRCC patients and drug repositioning for cancer treatment based on the analysis of shRNA- and drug-perturbed signature profiles of human kidney cell line.

Findings First, based on the gene co-expression network analysis, we identified two types of gene modules in ccRCC, which significantly enriched with unfavorable and favorable signatures indicating poor and good survival outcomes of patients, respectively. Then, we selected four genes, *BUB1B, RRM2, ASF1B* and *CCNB2*, as the potential drug targets based on the topology analysis of modules. Further, we repurposed three most effective drugs for each target by applying the proposed drug repositioning approach. Finally, we evaluated the effects of repurposed drugs using an *in vitro* model and observed that these drugs inhibited the protein levels of their corresponding target genes and cell viability.

Interpretation These findings proved the usefulness and efficiency of our approach to improve the drug repositioning researches for cancer treatment and precision medicine.

Funding This study was funded by Knut and Alice Wallenberg Foundation and Bash Biotech Inc., San Diego, CA, USA.

Keywords: Systems biology; Co-expression network; Target chemotherapy; Drug repositioning; ccRCC

Abbreviations: ccRCC, Clear cell renal cell carcinoma; CMap, ConnectivityMap; GCN, Gene co-expression network; OS, Overall survival; DEGs, Differentially expressed genes

*Corresponding authors at: Science for Life Laboratory, KTH-Royal Institute of Technology, Stockholm SE-17165, Sweden.
E-mail addresses: cheng.zhang@scilifelab.se (C. Zhang), adilm@scilifelab.se (A. Mardinoglu).

Research in context

Evidence before this study

Profile-based drug repositioning approaches have now been widely employed to predict the new disease-drug associations, which repurpose the well-characterized drugs for disease treatment and dramatically decrease the cost and duration taken by traditional drug development. The highly molecular heterogeneity, low response rates and drug resistance of ccRCC exacerbate the challenges in the tumor therapy. Therefore, it is needed to discover new treatment options for patients using the novel profile-based drug repositioning approaches.

Added value of this study

In this study, we found a set of targetable hub genes which controlled the ccRCC development were associated with the cell cycle dysfunctions. Further, we developed a novel drug repositioning approach based on the analysis of shRNA- and drug-perturbed transcriptomics profiles of cell line model and identified potentially effective drugs for each target gene. Importantly, we validated the drug efficacy in *in vitro* model, which provided new chance for the tumor treatment.

Implications of all the available evidence

We demonstrated the feasibility of the integrated approach combining the disease-target and drug-target prediction in the treatment of kidney cancer. Besides that, this approach could be also broadly applied to other cancers, which provides new insight into cancer treatment and precision medicine.

Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common histological subtype of renal cell carcinoma (RCC), accounts for 70% of all RCC cases.[1] Surgery (radical or partial nephrectomy) is the standard primary treatment for patients with localized tumors. The first-line and second-line target therapy options for patients with relapsed after nephrectomy or advanced stage tumor include tyrosine kinase inhibitors (axitinib, sorafenib, pazopanib, and sunitinib, etc.), mTOR inhibitors (everolimus and tesirolimus), and monoclonal antibodies against VEGF, PD-1 or PD-L1 (bevacizumab, pembrolizumab and avelumab, etc.). However, the National Comprehensive Cancer Network (NCCN, version: 1.2022) has reported that the response rates of the single-agent or combinatory regimens based on these drugs range from 6% to 50% in different clinical trials.[2] Moreover, the average duration of disease control with these drugs is only 8-9 months for the first-line setting and 5-6 months for the second-line setting.[3] Therefore, there is a need to discover more tolerated and effective drugs to widen the options for single-agent or combinatory regimens for ccRCC patients.

Computational drug repositioning based on systems biology methods has become a powerful tool to identify potential drug-target interactions and drug-disease interactions.[4] The advantage of drug repositioning is that the pharmacology and safety of the repositioned drugs have been well-characterized, dramatically decreasing the cost and duration taken by traditional drug development and reducing the risk of attrition in clinical phases.[5,6] In general, current drug repositioning strategies can be classified into drug-based, disease-based and profile-based.[7] Usually, drug-based and disease-based approaches are conducted by comparing drug-drug or disease-disease similarity or applying existing drug treatment knowledge to predict new disease-drug associations.[8,9] In contrast, profile-based approaches are conducted by analyzing the high-throughput multi-omics data associated with diseases and drugs, which do not rely on prior knowledge about a particular drug or disease and have increased ability to discover new drug-disease pairs.[7]

Recently, several studies have employed profile-based repositioning methods to identify potentially valuable drugs for the treatment of ccRCC. A widely used method is selecting the drug that has a reversed effect on the disease signature genes.[10,11] The idea of this method is that if the perturbation of gene expression induced by a drug (drug-perturbed signatures) is negatively correlated with the dysregulation in the tumor tissues compared to normal tissues (disease-specific signatures), this drug turns out to have therapeutic value for this tumor type. During the application of the above approach, ConnectivityMap (CMap)[12] is the most commonly used drug-perturbed gene expression data source, and it has been recently updated and integrated into the LINCS Data Portal.[13] To date, the LINCS data portal includes more than three million gene expression profiles associated with more than 20,000 drugs, gene overexpression and knockdown in up to more than 200 cell lines.[13] However, a drug repurposed in this way is supposed to have multiple gene targets mixed by oncogenes and passenger genes, limiting the identification and validation of the key targets and mechanisms of drug effect. To avoid this problem, we had a different strategy by starting from the target prediction and then repurposing the existed inhibitor of the target genes for cancer treatment. In our previous study, we identified a list of candidate target genes which are essential for the ccRCC tumor cell growth by the genome-scale metabolic model analysis.[14] Among these essential genes, we further filtered out three genes, *SOAT1, CRLS1* and *ACACB*, whose inhibition is not toxic in the 32 major normal human tissues by performing an *in silico* toxicity test for each essential gene, as the final targetable genes. Finally, we repurposed a well-known inhibitor of

SOAT1, mitotane, and validated its drug efficacy for the treatment of ccRCC. However, one limitation of this study is that a systematic drug repositioning approach is needed to discover more potentially effective drugs that could inhibit *SOAT1*. In addition, the candidate target genes are also needed to extend, which can be not only limited to the metabolically important genes that only covers around 20−25% of human genome as reported in the human metabolic model[15] but also can be non-metabolic genes. In contrast, gene co-expression network (GCN) analysis that applies all possible human genes has been widely used to identify the key genes, its neighbors and functionally related biological functions.[5,16]

In this study, we applied an integrated strategy that involved disease-target prediction and drug-target prediction. First, we extracted a set of robust ccRCC signature genes whose expression was significantly associated with patients' survival outcomes. Functional enrichment analysis showed that these genes were significantly enriched in the cell division, cell cycle, DNA replication, angiogenesis, cell migration and cell differentiation pathways, all of which are well-known hallmarks in cancer.[17] Second, we identified two types of molecular modules that significantly enriched with the unfavorable and favorable signatures based on the GCN analysis of the transcriptomic data of ccRCC tissues. Third, we extracted four target genes, including *BUB1B, RRM2, ASF1B* and *CCNB2,* showing high centrality in the modules. Next, we developed a drug repositioning approach based on the analysis of shRNA-perturbed and drug-perturbed transcriptomics data from the LINCS data portal and repurposed the three most effective drugs for each target. Finally, we tested the drug effect in Caki-1 cells and showed the efficacy of the selected drugs Figure 1. showed the whole study design.

Methods

Data and preprocessing

Transcript-expression profiles (TPM and count values) of 528 TCGA ccRCC samples and 72 adjacent normal samples were downloaded from https://osf.io/gqrz9.[18] We extracted the tumor and normal samples with sample and vial identifiers of BRC patient barcodes '01A' and '11A', respectively, which represent primary solid tumor tissue and solid normal tissue from the first vial, respectively. The mRNA expression was quantified by Kallisto[19] based on the GENCODE reference transcriptome (version 24) (Ensembl 83 (GRCh38.P5)). The clinical information of TCGA samples was downloaded by using the R package TCGAbiolinks.[20] The mRNA-seq data of 100 ccRCC samples of patients from the Japanese cohort[21] was downloaded from the European Genome-phenome Archive (Accession number: EGAS00001000509). BEDTools[22] was used to convert BAM to FASTQ file, and Kallisto was used for quantifying the count and TPM values of transcripts based on the same reference transcriptome of TCGA data. The sum value of the multiple transcripts of a gene was used as the expression value of this gene. The genes with average TPM values >1 were analyzed.

The essential scores of genes of 16 ccRCC cell lines were downloaded from DepMap Portal (https://depmap.org/portal/),[23] which are estimated based on the CRISPR-Cas9 essentiality screens. The meaning of the score is the essentiality of a gene for cancer cell survival after CRISPR-Cas9 knockout of this gene. More negative scores indicate more essential. The processed RNA-seq data of Caki-1 was downloaded from the Cancer Cell Line Encyclopedia (CCLE) portal (version: CCLE_RNA-seq_rsem_transcripts_tpm_20180929.txt.gz).[24]

Ethics

Since all the human RNA-seq data were downloaded from the public databases, there was no ethical issue.

Statistics

Both the univariate Cox regression model and the Kaplan-Meier method were used to evaluate the association of gene expression with patients' overall survival (OS). For Cox analysis, we used a R package named 'survival' in which the input data is the expression values of each gene (TPM values) across all samples of patients, OS and survival outcomes of patients (death or alive). For Kaplan-Meier analysis, we classified the patients into two groups based on the TPM values of each gene and examined their prognoses. Survival curves were estimated by the Kaplan-Meier method and compared by log-rank test. To choose the best TPM cutoffs for grouping the patients most significantly, all TPM values from the 20th to 80th percentiles used to group the patients, significant differences in the survival outcomes of the two groups were examined, and the value yielding the lowest log-rank P-value was selected.

DESeq2[25] was used to identify differentially expressed genes (DEGs) between TCGA tumor tissues and normal tissues groups. The lowly expressed genes with average TPM≤1 were removed, and the raw count values of the remaining genes were used as the input of DESeq2. False discovery rate (FDR) was adjusted by the Benjamini-Hochberg (BH) method. FDR<0.05 was used to identify significant DEGs.

Functional enrichment analysis

The enrichGo function of the R package ClusterProfiler was used for Gene Ontology (GO) enrichment,[26] which uses the hypergeometric distribution to estimate whether a list of genes is significantly enriched in each

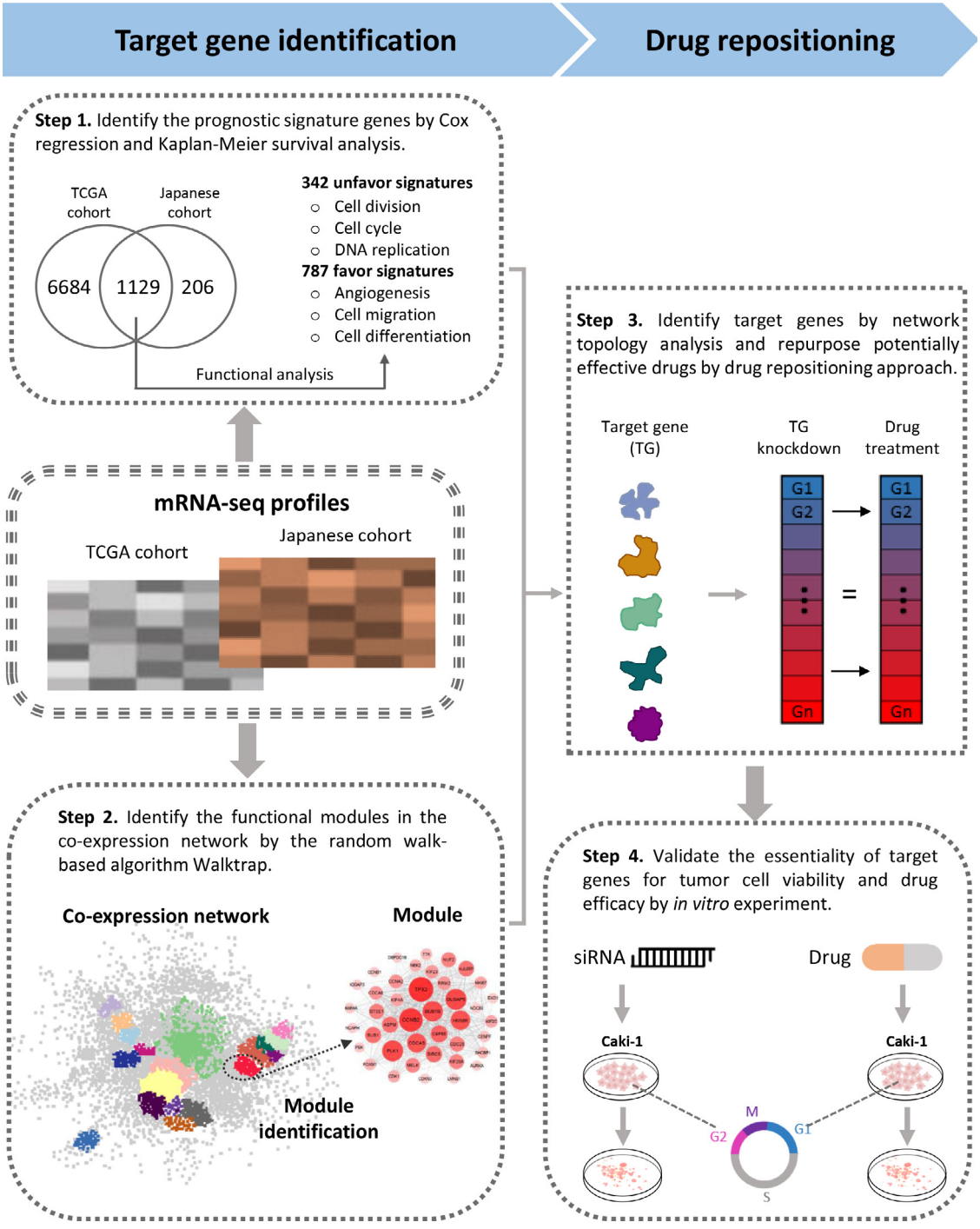

Figure 1. The flowchart for the whole study design including target gene identification and drug repositioning for target genes. Step 1: we identified a set of robust ccRCC prognostic signature genes based on the gene expression profiles of two independent ccRCC cohorts and the expression of these signature genes could indicate the survival outcomes of ccRCC patients. Step 2: we conducted the GCN analysis in each of the two cohorts and identified two types of reliable gene modules which significantly enriched with unfavorable and favorable signature genes, respectively. Step 3: several hub genes with consistently high centralities in the modules were selected as druggable target genes based on the network topology analysis. Further, we developed a drug repositioning approach based on a systematic study of shRNA- and drug- perturbed signature profiles from CMap. Based on this approach, we repurposed the three most effective drugs for each target. Step 4: we evaluated the essentiality of target genes and inhibitory effects of repurposed drugs in a ccRCC cell model Caki-1.

GO pathway. FDR<0.05 was used to identify significantly enriched pathways.

GCN analysis and gene module identification

Spearman correlation was used to estimate the correlation between each two genes across all tumor samples. The gene-gene links ranked in the top 1% with the highest correlation coefficients were extracted to construct the co-expression networks. A random walks-based algorithm, Walktrap,[27] was used to capture the gene modules with high transitivity from the co-expression network. The modules with more than 20 nodes and clustering coefficients higher than 0.6 were extracted for further analysis. R package 'igraph' was used for the topology analysis.[28]

Concordance analysis of the prognostic genes

If two lists of prognostic genes, list 1 with L_1 genes and list 2 with L_2 genes, have k overlapping genes, among which s genes show the exact prognostic directions (both favorable or unfavorable) in the two gene lists, the probability of observing at least s consistent genes by chance can be estimated based on the following cumulative hypergeometric distribution model:

$$P = 1 - \sum_{i=0}^{s-1} \frac{\binom{L_2}{i}\binom{L-L_2}{L_1-i}}{\binom{L}{L_1}}$$

Where, L represents the number of the background genes commonly measured in the datasets from which the prognostic genes are extracted. The two gene lists were considered to be significantly overlapped if $P < 0.05$. The concordance score s/k is used to represent the consistency between the two lists of genes. The score ranges from 0 to 1, and the higher concordance score indicates the better consistency of two lists of genes.

The hypergeometric distribution model was also used to test whether the favorable or unfavorable signatures significantly overlap with the genes involved in a functional module.

Drug repositioning

The shRNA- and drug-perturbed signature profiles (level 5 data) in HA1E, a kidney cell, were downloaded from the CMap data portal (https://clue.io/, version: CMAP LINCS 2020).[12] Three or more biological replicates typically do the experiments in CMap. The level 5 data provides the replicate-collapsed Z-scores, representing a consensus biological response of transcriptomics to the perturbation of drug treatment or shRNA infection derived from different replicates.[29] Totally, we obtained 37,669 drug-perturbed signature profiles related to 6,986 drugs with different doses and time points and 21 shRNA-perturbed signature profiles related to *BUB1B* (six shRNAs), *RRM2* (three shRNAs), *CEP55* (three shRNAs), *ASF1B* (three shRNAs) and *CCNB2* (six shRNAs).

Our drug repositioning approach hypothesizes that a drug is considered to have an inhibitory effect on the expression of a target gene if this drug leads to a wide perturbation on the gene expression landscape in tumor cells, similar to the effect of the knockdown of the target gene. We applied four procedures to identify the drugs which had the highest possibility to inhibit the expression of their corresponding target genes by an integrated analysis of the shRNA- and drug-perturbed signature profiles (see Results section and Figure 5a for details). (1) Constructing the drug-shRNA matrix for each target gene. For a given target gene, we extracted its shRNA-perturbed signature profiles in HA1E cell line in which the Z-scores represent the biological dysregulation of gene expression in cells after shRNA infection. As one gene was knocked down by at least three shRNAs in the experiment setting in CMap, the shRNA-perturbed signature profiles were presented as a gene-shRNA matrix. We also extracted the drug-perturbed signature profiles in HA1E cell line in which the Z-scores represent the biological dysregulation of gene expression in cells after drug treatment and thus generated a gene-drug matrix. Then, we calculated the Spearman correlation between each two possible lists of drug-perturbed and shRNA-perturbed signatures and thus generated a correlation coefficient matrix, named drug-shRNA matrix, in which each row represents a drug treatment by a specific dose and time point, and each column represents a specific shRNA for the knockdown of this target gene. A correlation coefficient in the drug-shRNA matrix represents the similarity of effects on gene expression between specific drug treatment and specific shRNA knockdown. (2) Optimizing the drug-shRNA matrix. Since the drug treatment with different doses and time points could induce different effects on gene expression in cells as well as different similarities with the effects induced by shRNA knockdown, we only extracted the optimal dose and time point for each drug whose perturbation showed the highest similarity (correlation coefficient) with shRNA infection. Thus, we simplified the rows by keeping a unique and optimal dose and time point for each drug in the drug-shRNA matrix. A cell line may respond differently to different shRNAs even though they target the same genes. Thus, we simplified the column by extracting the optimal shRNAs whose signatures consistently correlate with drug perturbed signatures in the drug-shRNA matrix. We performed a clustering analysis of different shRNAs based on the correlation of different columns in the drug-shRNA matrix and extracted the shRNAs, which were clustered together and showed better similarity with drug-perturbed effects. Finally, we extracted three shRNAs for *BUB1B*, two shRNAs for *RRM2*, two shRNAs for *CEP55*, two shRNAs for *ASF1B*, and four shRNAs for *CCNB2*. (3) Extracting the top 1% drug

candidates. Based on the optimized drug-shRNA matrix, we ranked the drugs associated with each shRNA based on the correlation coefficient from the largest to smallest. A higher rank indicated a higher correlation. We first extracted a list of drugs involved in the top 1% with the highest ranks associated with each shRNA, and then the overlapped drugs among the top 1% lists related to different shRNAs were selected as candidates. (4) Selecting the three most effective drugs for each target. The top three drugs with the highest median ranks were finally considered the most effective drugs for each target.

We took the drug repositioning for *BUB1B* as an example. The drug-shRNA matrix of *BUB1B* included 37,669 rows associated with 6,986 drugs with different doses and time points and six columns related to six different shRNAs. After optimization, we obtained a simplified drug-matrix with 6,986 drugs, each with the best dose and time point and three optimal shRNAs for knockdown of *BUB1B*. Then, we extracted the top 1% drugs (70 drugs) with the highest correlation coefficients associated with each shRNA. 24 overlapped drugs were found among the three lists of top 1% drugs. Among these drugs, the top three drugs, TG-101209, oxetane, and WH-4-025, with the highest median ranks, were finally selected as the most effective drugs for targeting *BUB1B*.

IHC image

The IHC images of ccRCC tumor tissues and normal kidney tissues were downloaded from the Human Protein Atlas website (https://www.proteinatlas.org/).[30,31] The protocol of IHC was provided in.[30,31] For a given gene, we selected the IHC images of normal tissue and tumor tissue conducted by the same antibody. *TPX2, ASF1B, CCNB2* and *TCF4* IHC images were done by the HPA005487, HPA069385, HPA008873 and HPA020722 antibodies, respectively.

Cell culture

Human ccRCC cell line Caki-1 was obtained from the Karolinska Institute of Environmental Medicine, Stockholm, Sweden, derived from a male ccRCC patient. The cells were cultured in RPMI 1640 medium with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. The cells were cultivated at 37°C in a humidified incubator with 5% CO_2.

siRNA transfection

20×10^3 cells were plated by quadruplicate into a 96-well plate per well. Day after seeding, 1 pmol siRNAs for targeting *BUB1B, RRM2, ASF1B, CCNB2* and *CEP55* (OriGene Technologies Inc., USA) were transfected into cells using the Lipofectamine® RNAiMAX reagent (Invitrogen) for three days. Then, we used the cells for cell viability assay and harvested protein lysate for Western blot analysis after 3-day transfection.

Drug treatment

The drugs, TG-101209, NVP-TAE684, MK-0752, actinomycin-d, and panobinostat were purchased from Selleckchem (S2692, S1108, S2660, S8964, S1030, Selleckchem, Houston, TX, USA), and withaferin-a was purchased from Sigma (W4394, Sigma-Aldrich, Saint Louis, MO, USA). These drugs were dissolved in DMSO. The cells were seeded in a 96-well plate at 20×10^3 cells per well. Day after seeding, drugs were added to the wells at a proper concentration, TG-101209 (6nM), NVP-TAE684 (3nM), MK-0752 (5nM), withaferin-a (4uM), actinomycin-d (0.3nM), and panobinostat (5nM) and treated for one day. The cells in the negative control group were only treated by DMSO for one day.

Western blots

The cells were washed with PBS and lysed with CelLytic M (C2978, Sigma-Aldrich, Saint Louis, MO, USA) lysis buffer containing protease inhibitors. The lysates were centrifuged at 12,000 rpm for 5 min, and the supernatant was collected. The proteins were separated by Mini-PROTEAN® TGXTM Precast Gels (BioRad, Berkeley, CA, USA) and transferred to a Trans-Blot Turbo Mini 0.2um PVDF Transfer Packs membrane (BioRad, Berkeley, CA, USA) by using Trans-Blot® TurboTM Transfer System (Bio-Rad, Berkeley, CA, USA). The antibodies for BUB1B (HPA008419), RRM2 (HPA056994), CEP55 (HPA023430), ASF1B (HPA069385), CCNB2 (HPA008873), and GAPDH (sc47724, Santa Cruz Biotechnology, Inc.) were used for primary immunoblotting. All the antibodies were diluted at 1:10000 concentration. The membranes were incubated in primary antibody solution overnight at 4°C with gentle rocking. Secondary antibody, goat Anti-Rabbit HRP (ab205718) or goat anti-mouse IgG-HRP (sc2005, Santa Cruz Biotechnology, Inc.), was blotted for 30 min at 4°C with gentle rocking. The protein bands were detected with ImageQuanattm LAS 500 (29-0050-63, GE) automatic exposure procedure or 6 min exposure.

Cell viability assay

Cell proliferation was detected by Cell Counting Kit-8 (CCK-8) assay. The 10 ul of CCK-8 reagent (1:10) was added to each well of 96-well plate with siRNA transfected cells or drug-treated cells by manufacturer's instruction. The 96-well plate was incubated at 37°C for 2 h and then measured absorbance at 450 nm using a microplate reader (Hidex Sense Beta Plus).

Role of the funding source

This study was funded by Knut and Alice Wallenberg Foundation and Bash Biotech Inc., San Diego, CA, USA. The funder has no role in study design, data collection, analysis, interpretation and writing of the report. The corresponding author had full access to all the data in this study and held the final responsibility for the decision to submit for publication.

Results

Identification of ccRCC signature genes

We applied both Kaplan-Meier analysis and univariate Cox model to investigate the associations of the mRNA expression levels of genes with the patients' OS in TCGA and Japanese ccRCC cohorts. We performed a Kaplan-Meier survival analysis by classifying the patients into two groups with high and low expression (TPM values) of the investigated gene by selecting an optimal cut off from the 20th and 80th expression percentiles yielding the lowest log-rank P-value as in our previous study.[32,33] Meanwhile, we performed the univariate Cox analysis by calculating the hazard ratio of each gene. As a result, we identified an overlap of 7,813 prognostic genes by Kaplan-Meier and Cox survival analysis in the TCGA cohort (FDR<0.05). Similarly, we identified 1,335 prognostic genes in the Japanese cohort (FDR<0.05). The two sets of genes had a significant overlap (n=1,129, hypergeometric distribution test, P < 1.11e-16), and the concordance score of these overlapped genes (both favorable or unfavorable genes) is 99.91% (Figure 2a). Among these 1,129 prognostic genes, there were 342 unfavorable genes and 787 favorable genes whose high expression indicated poor and good survival outcomes of patients, respectively. Functional enrichment analysis showed that the unfavorable genes were significantly enriched in cell division and cell cycle pathways. In contrast, the favorable genes were significantly enriched in the angiogenesis, vasculogenesis, cell migration and cell differentiation pathways (FDR<0.05, Figure 2b), which are well-known hallmarks in cancer.[17] Therefore, we used these genes as unfavorable and favorable signature genes for ccRCC.

Identification of functional modules in the GCN

GCN analysis is a powerful method to identify the hub genes that drive the key cellular signaling pathways, including a set of co-expressed or functionally associated genes during tumor development. The conservative co-expressed pattern may confer a selective advantage for tumor cells if it could be validated in independent datasets.[34,35] We calculated the Spearman correlation between two possible gene pairs based on the mRNA expression profiles of ccRCC patients in TCGA and Japanese cohort, respectively. To balance the sensitivity for detecting the functional gene modules and the robustness for validation in independent datasets, gene-gene links involved in the top 1% with the highest correlation coefficients were extracted to build the GCN. We obtained around 900,000 gene-gene links with correlation coefficients ranging from 0.74 to 1 in the co-expression network of the TCGA cohort. Then, we used a random walks-based algorithm, Walktrap,[27] to identify the gene modules with high transitivity based on the topology of the GCN. To discover functionally meaningful modules in which genes show a good connectivity, we extracted the modules with more than 20 genes and clustering coefficients higher than 0.6. Finally, we obtained 18 modules (M1-18) in the TCGA cohort (Figure 2c). Further, we investigated the association of each module with the unfavorable and favorable signature genes by concordance analysis (Table S1). We found that the 41 genes involved in M11 had a significant overlap with the 342 unfavorable signature genes (n=40, hypergeometric test, FDR<0.05). Thus, we determined the M11 as an unfavorable module. Meanwhile, we observed that the genes involved in M1 (184 genes), M2 (177 genes), M3 (88 genes), M4 (76 genes) and M18 (22 genes) had significant overlaps with the 787 favorable signature genes, respectively (n=51, 86, 33, 18 and 6, hypergeometric test, FDR<0.05), which were determined as favorable modules. Then, we investigated the biological function of these modules by functional enrichment analysis. We found that the genes involved in unfavorable module were significantly enriched in cell division, cell cycle, DNA replication and MAPK activity pathways (FDR<0.05, Figure 2c). The genes involved in favorable modules were significantly enriched in the angiogenesis, vasculogenesis, cell migration, cell differentiation, cell adhesion and several signaling pathways (cAMP-mediated signaling, ERK1 and ERK2 cascade, Ras protein signal transduction and Rho protein signal transduction) (FDR<0.05, Figure 2c).

Based on the same methodology, we obtained around 900,000 gene-gene links with correlation coefficients ranging from 0.68 to 1 in the GCN of the Japanese cohort. We found five modules (M1-5) based on the exact cutoff setting (Figure 2d). M3 (81 genes) and M1 (770 genes) had significant overlaps with the unfavorable and favorable signature genes, respectively (n=76 and 222, hypergeometric test, FDR<0.05, Figure 2d, Table S2), which were correspondingly determined as unfavorable and favorable modules. Functional enrichment analysis of the two modules exhibited an identical set of biological pathways resulting from the TCGA cohort (Figure 2d). Based on these results, we inferred that M11 from the TCGA cohort and M3 from the Japanese cohort indicated the same unfavorable gene module and M1, M2, M3, M4 and M18 from the TCGA cohort and M1 from the Japanese cohort indicated the same favorable gene module. As shown in Figure 2e

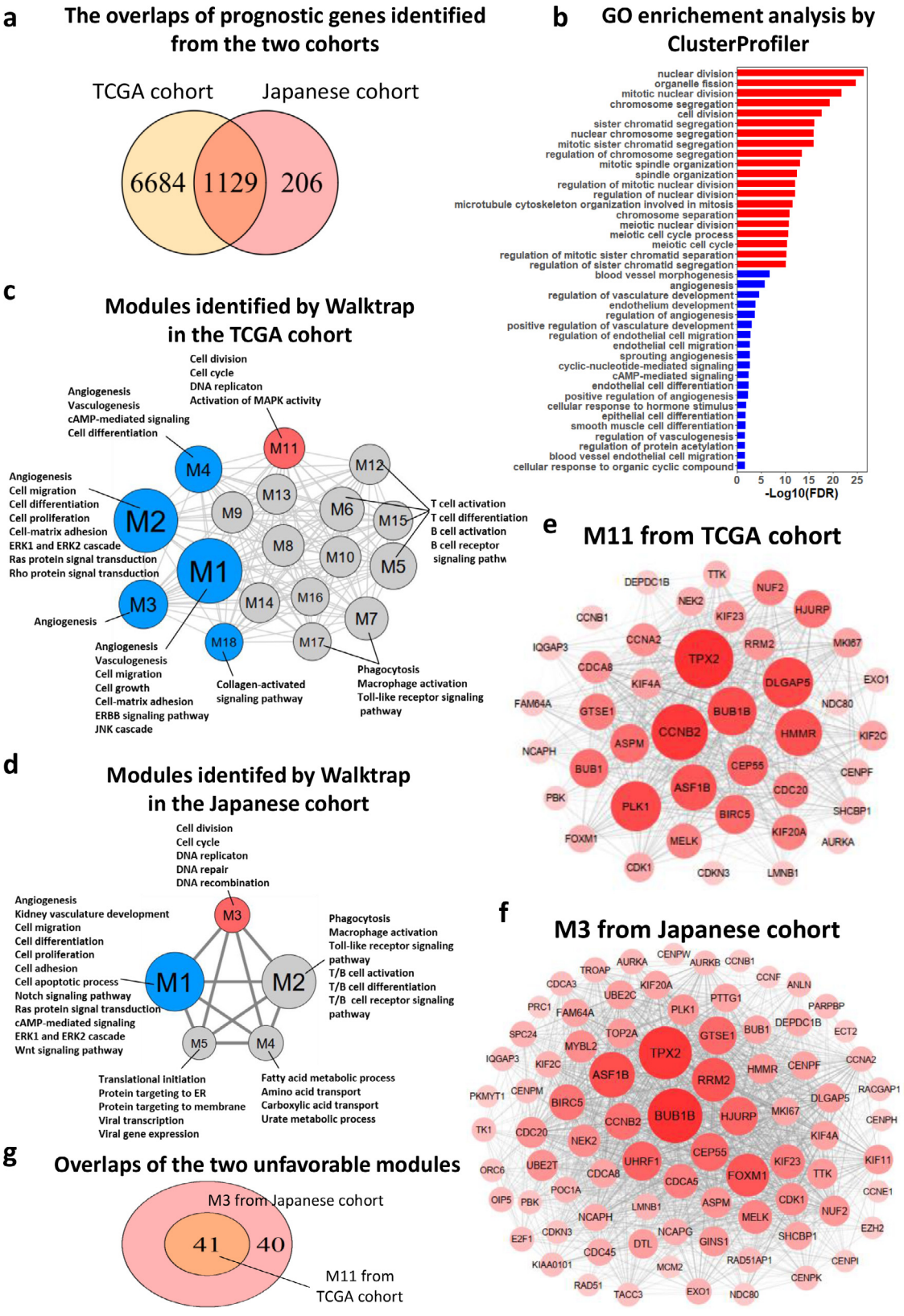

Figure 2. Identification of signature genes and functional gene modules for ccRCC. (a) Venn diagram showing the consistency of prognostic genes identified from TCGA and Japanese cohorts with 528 and 100 patients, respectively. (b) Top 20 most significantly

and 2f, we visualized the individual genes in these two modules. The M3 from the Japanese cohort wholly covered all the genes involved M11 from the TCGA cohort, which is highly significant overlap (n=41, hypergeometric test, P < 1.1e-16, Figure 2g). Moreover, each of M1, M2, M3, M4 and M18 from the TCGA cohort had a significant overlap with M1 from the Japanese cohort (n=98, 168, 78, 73 and 20, hypergeometric, all P < 1.1e-16) (Table S3). Thus, we merged the genes from M1, M2, M3, M4 and M18 in the TCGA cohort and generated an integrated module M1/2/3/4/18. These findings suggested that we found two types of gene modules that lead to unfavorable and favorable survival outcomes of patients in ccRCC, which had high confidence since these modules were validated in an independent cohort with different racial and geographical characteristics.

We found the genes involved in M7 and M17 from the TCGA cohort were significantly enriched in the innate immune response pathways (phagocytosis, macrophage activation and Toll-like receptor signaling pathway) and the genes involved in M5, M6, M12 and M15 were significantly enriched in the adaptive immune response pathways (T/B cell activation, T cell differentiation and B cell receptor signaling pathway). Interestingly, the genes involved in M2 from the Japanese cohort were significantly enriched in an identical set of both innate and adaptive immune response pathways. Concordance analysis showed that each of the modules from the TCGA cohort mentioned above had a significant overlap with M2 from the Japanese cohort (Table S4), suggesting these immune-related modules identified from the TCGA cohort were also independently validated in the Japanese cohort. In addition, the genes involved in M4 from the Japanese cohort were enriched in the fatty acid metabolic process. In our previous study, we have reported three molecular subtypes of ccRCC characterized by the different activity of energy metabolism in terms of cell viability and proliferation, cell differentiation and fatty acid metabolism,[14] which are consistent with the functions of the modules we identified in this study. The genes involved in M5 from the Japanese cohort were significantly enriched in the translational initiation, protein targeting to the endoplasmic reticulum, protein targeting to membrane, viral transcription and viral gene expression pathways. Interestingly, two of the three subtypes we identified in our previous study were characterized by the opposite dysregulation of these pathways, corresponding to patients' best and poorest survival outcomes.[14]

Identification of the potentially druggable target genes

Analysis of network topology could provide insight into the importance of a given gene in the network. The genes with high centrality are considered as hub genes in the network, and we predicted that these are potential targets that are worthwhile to be further evaluated. We calculated the degree, betweenness and closeness of each gene in each unfavourable and favourable module. These three measurements are commonly used to characterize the centrality of nodes from three distinct perspectives: a higher degree indicates that the node is involved in more interactions; a higher betweenness indicates that the node acts as a bridge which lies on the shortest path between other nodes; a higher closeness indicates that the node shows shorter paths to all the other nodes and is likely to be the geometric center of the module.[36–39] We then calculated the Spearman correlation of each centrality measurement between the two unfavorable modules (or favorable modules) identified from TCGA and Japanese datasets. As shown in Figure 3a and 3b, the correlation coefficients of degree, betweenness and closeness were 0.77, 0.73 and 0.78 between M11 from the TCGA cohort and M3 from the Japanese cohort, and 0.73, 0.71 and 0.72 between M1/2/3/4/18 from TCGA cohort and M1 from the Japanese cohort. This result suggests that the topology of the unfavorable and favorable modules identified from the TCGA cohort are highly similar to the Japanese cohort even though the two cohorts have different racial and geographical characteristics. We ranked the genes involved in each unfavourable and favourable module based on decreasing order of the degree, betweenness and closeness values. We then selected the common genes involved in the top 10 gene lists with the highest rank based on at least one of the three centrality measurements in both TCGA and Japanese cohorts as the hub genes. Consequently, we found six hub genes, namely *BUB1B, TPX2, RRM2, CEP55, ASF1B* and *CCNB2*, from the two unfavorable modules (Figure. 3c and S1a) and four hub genes, including *ERG, ARAP3, TCF4* and *MMRN2*, from the two favorable modules (Figure. 3d and S1b). Here we ignored *CD93* in the

GO pathways enriched with unfavorable (red bars) and favorable signature genes (blue bars), respectively. The gene modules identified in (c) TCGA and (d) Japanese cohorts. Only the modules with more than 20 genes and clustering coefficients higher than 0.6 are shown. Two modules are linked if there are a positive correlation between the genes from these two modules. Sizes of modules are correlated with the corresponding numbers of genes from the modules. The red nodes indicate the modules significantly enriched with unfavorable signature genes. The blue nodes indicate the modules significantly enriched with favorable signature genes. ER, endoplasmic reticulum. Visualization of the genes involved in (e) M11 from TCGA cohort and (f) M3 from Japanese cohort. The sizes and color of nodes are correlated with the betweenness values of genes in the modules. (g) Venn diagram showing the consistency of genes involved in the M11 from TCGA cohort and M3 from Japanese cohort.

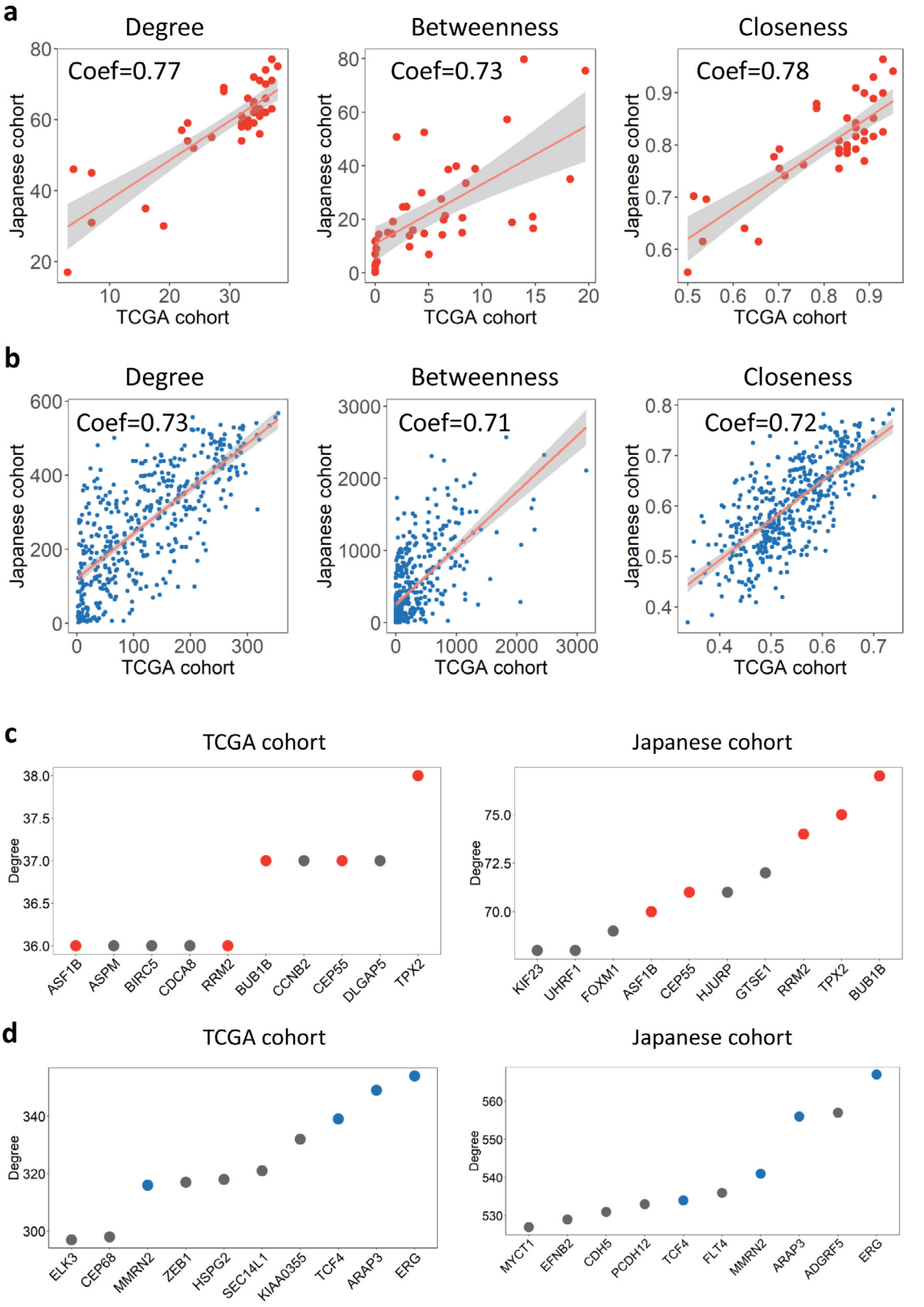

Figure 3. Topology analysis of the unfavorable and favorable modules. (a) Spearman correlation of degree, betweenness or closeness centralities of genes from M11 in the TCGA cohort and M3 in the Japanese cohort. Coef, Spearman correlation coefficient. (b)

favorable modules since it had no significant association with prognoses of patients. The prognostic effects of these hub genes were shown in Figs. S2 and S3.

We visualized the expression levels of the 10 hub genes based on the gene expression profiles of TCGA tumors and adjacent normal samples. As shown in Figure 4a and 4b, most of these hub genes, especially the six unfavorable hub genes, presented very low expression in normal tissues, while their expression levels were significantly increased in the tumor tissues. As these unfavorable hub genes encode cell mitosis and cell cycle-related proteins, this alteration may indicate that low expression of these genes maintains normal cell renewal and metabolism, but increased expression promotes tumorigenesis or tumor progression. Meanwhile, the significantly increased expression levels of favorable hub genes in tumor tissue may imply activating a protective mechanism to mediate anti-tumor effect in cancer. Further, we investigated the altered coverage of these hub genes in TCGA tumor tissues. The result showed 84-96% of patients presented a high expression of unfavorable hub genes, which leads to poor survival outcomes of patients and 59-73% of patients showed an increased expression of favorable genes, which leads to good survival of patients (Figure 4c), suggesting that most of the ccRCC patients may carry the common molecular alterations induced by the unfavorable hub genes but not for their anti-tumor mechanisms. To confirm the expression changes at the protein level, we validated the expression differences that we observed at the mRNA level using the immunohistochemistry (IHC) images of the hub genes in normal tissue and tumor tissue from the Human Protein Atlas[30,31] (Figure 4d). Four hub genes with selected antibodies showed a higher protein expression in tumor cells than corresponding normal tissues. For example, TPX2 was not detected in normal glomeruli cells, and lowly expressed in normal tubule cell, but showed medium expression in tumor cells (Figure 4d).

Moreover, we extracted the essential scores of these hub genes in 16 ccRCC cell lines from the Dependency Map (DepMap) portal.[23] The essential scores were evaluated based on genome-scale CRISPR-Cas9 loss-of-function screens and corrected by a computational method CERES.[23] A more negative score of a gene indicates this gene is more essential for tumor cell proliferation and survival. As shown in Figure 4e, most of the unfavorable hub genes were more essential than the favorable genes in ccRCC. We also investigated the mRNA expression levels of these hub genes in Caki-1, a commonly used ccRCC cell model. The result showed that the unfavorable hub genes had appropriate expression levels in Caki-1 (Figure 4f). Considering these, the unfavorable hub genes are considered more appropriate target genes, and the inhibitory effect on these genes (e.g., shRNAs and chemical inhibitors) should be further evaluated.

Drug repositioning for target genes

To find the potential therapeutic drugs for candidate targets, we developed a drug repositioning approach based on an integrated analysis of the drug- and shRNA-perturbed signature profiles from the CMap database.[29] We hypothesized that a drug has an inhibited effect on the expression of a target gene in tumor cells if the drug treatment leads to a similar dysregulation of gene expression induced by shRNA knockdown. The outline of our drug repositioning approach was described in Figure 5a. In brief, it consists of the following four steps (see Method section for details): (1) Constructing the drug-shRNA matrix for each target gene. We extracted the drug-perturbed and shRNA-perturbed signature matrix of the HA1E cell line, the only kidney cell line in CMap database. *TPX2* was not analyzed in the following analysis because its shRNA knockdown data was not provided in CMap. For each of the other five target genes, we constructed a correlation matrix, namely drug-shRNA matrix, by calculating Spearman correlation between all possible combinations of drug-perturbed and target-specific shRNA-perturbed signatures. The correlation coefficients in this matrix represent the similarity of effects on gene expression induced by specific drug treatment and specific shRNA knockdown. (2) Optimizing the drug-shRNA matrix. For a specific drug, different doses and treated time points were set for cell line treatment in the CMap experiments. Thus, we usually obtained several perturbed profiles associated with the same drug treatment in a cell line. It has been reported that the effects of drug on human cells is highly dependent on the dose setting.[40] Under different doses, drug-target binding affinities could be changed and the downstream pathways can be affected differently, especially for cell cycle related pathways.[40] To maximally represent the drug efficacy, we only kept the optimal dose and time point with the highest similarity with each shRNA knockdown. For a specific gene knockdown experiment, different shRNAs were set for

Spearman correlation of degree, betweenness, and closeness centralities between genes from the merged module M1/2/3/4/18 in the TCGA cohort and M1 in the Japanese cohort. Coef, the Spearman correlation coefficient. The gray area indicates the 95% confidence interval for the prediction of the linear model. (c) The top 10 genes with the highest degree values in M11 from the TCGA cohort and M3 from the Japanese cohort, respectively. The common genes between the two cohorts were highlighted by red color. (d) The top 10 genes with the highest degree centrality in M1/2/3/4/18 from the TCGA cohort and M1 from the Japanese cohort, respectively. The common genes between the two cohorts were highlighted by blue color.

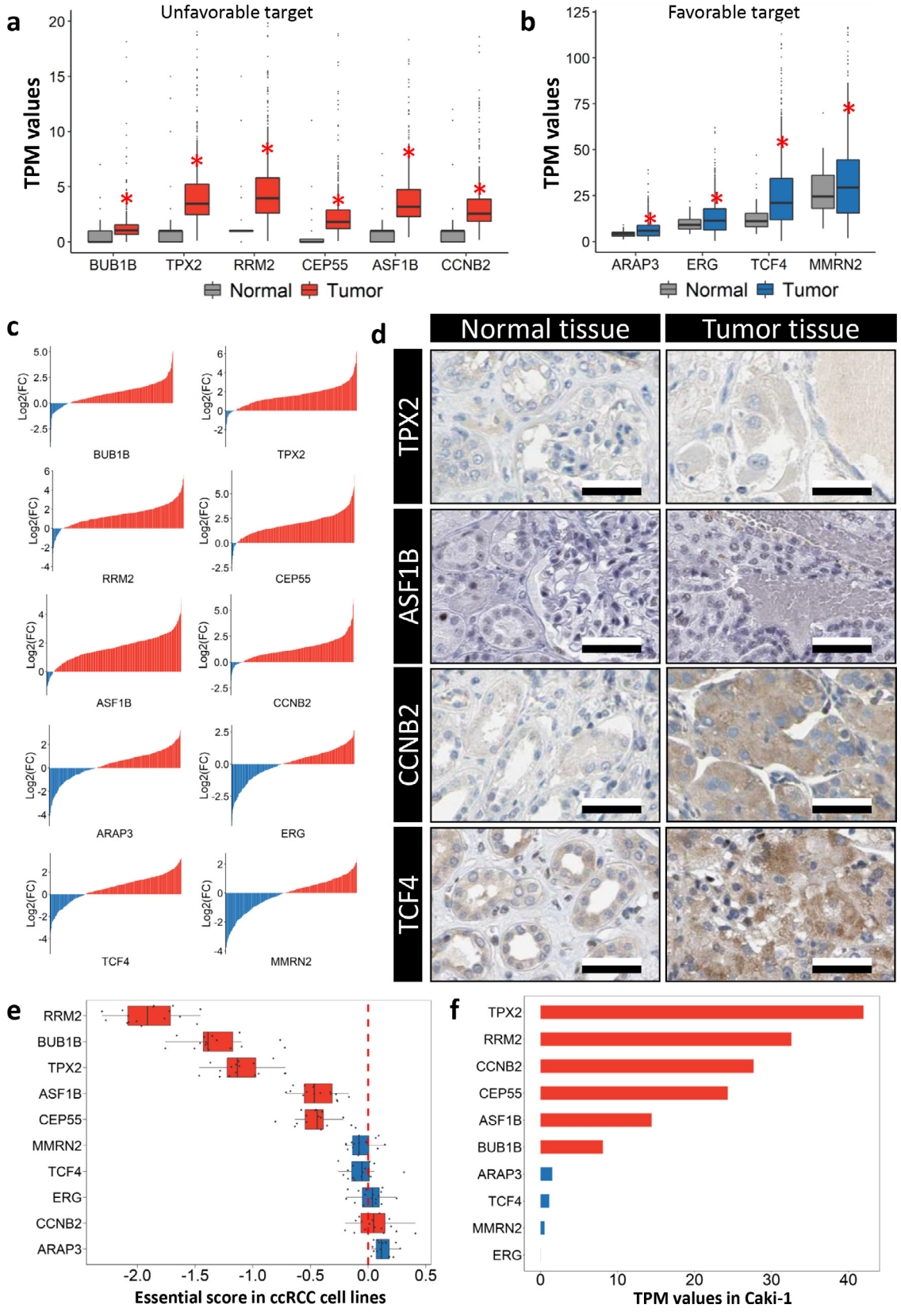

Figure 4. The alteration of hub genes in tumor tissues or cell lines. (a) Box plots showing the mRNA expression levels (TPM values) of unfavorable hub genes in TCGA 528 tumors compared to 72 normal tissues. The differential expression analysis was performed by

targeting the same gene. These shRNAs are the specific sub-sequences from the investigated gene, whose sequences could be completely different or share some common bases. To maximally represent the gene knockdown efficacy and avoid the effects from the off-target shRNAs, we extracted at least two shRNAs which showed a consistently higher correlation with drug treatment by a clustering analysis (Fig. S4). As a result, we obtained five simplified drug-shRNA matrixes, each of the matrixes had 6,986 drugs (rows) while three shRNAs (columns) for *BUB1B*, two shRNA for *RRM2*, two shRNA for *CEP55*, two shRNA for *ASF1B*, and four shRNA for *CCNB2*, respectively (Tables S5−S9). (3) Extracting the top 1% drug candidates. In a given drug-shRNA matrix, we ranked the drugs based on the correlation with each shRNA. We extracted the overlapped drugs, which ranked in the top 1% (70) drug lists with the highest correlation with each shRNA as candidates. (4) Selecting the three most effective drugs for each target. The top three drugs with the highest median rank across different shRNAs were finally considered as the most effective drugs for each target. As shown in Figure 5b, we discovered that TG-101209, oxetane, WH-4-025 target *BUB1B*, NVP-TAE684, MK-0752, and withaferin-a target *RRM2*, actinomycin-d, triphenyl-tin, RS-I-002-6 target *CEP55*, BRD-K26510616, panobinostat and tacedinaline target *ASF1B* and oxetane and BRD-K54343811 target *CCNB2*.

Validation of target genes and drug effect

To confirm the essentiality of the target genes in ccRCC, we inhibited the expression of target genes by transfecting siRNAs to Caki-1 and investigated the effect on cell viability. Western blots showed that the protein expression of the five target genes was successfully suppressed in Caki-1 transfected with siRNAs compared with negative control (Figure. 6a−e and S5). We performed a CCK-8 assay to measure cell proliferation. The result indicated that the cell viability was significantly decreased by the knockdown of *BUB1B, RRM2, ASF1B* and *CCNB2* in Caki-1 but not for the knockdown of *CEP55* (Figure 6a−e). Thus, *CEP55* was excluded in the following experiments.

Further, we investigated whether the drugs we repurposed could inhibit their corresponding target genes. Only TG-101209, NVP-TAE684, MK-0752, withaferin-a, actinomycin-d, and panobinostat were successfully purchased, and these drugs were tested in Caki-1 in the following experiments. As shown in Figure. 6f and S6a, we observed that the protein level of *BUB1B* was significantly decreased by the treatment of TG-101209, which was identified as the most effective drug for *BUB1B* compared to the negative control, and the cell viability was also significantly reduced. NVP-TAE684, MK-0752, and withaferin-a were the top three effective drugs that we identified for targeting *RRM2*. As shown in Figure. 6g and S6b, c, we observed the protein level of RRM2 was significantly suppressed by both NVA-TAE684 and withaferin-a but not by MK-0752. The cell viability was significantly reduced by NVA-TAE684 and withaferin-a but not by MK-0752, consistent with the result showing in Western blots. We also tested the effect of actinomycin-d, actually targeting *CEP55* based on our prediction, on RRM2. The Western blot showed that this drug does not work for inhibiting *RRM2*. The protein level of *ASF1B* was significantly decreased by its target drug panobinostat and cell viability was also significantly reduced after the drug treatment (Figure. 6h and S6d). Unfortunately, we could not test the effect of *CCNB2* since its target drugs are not available for purchase. In summary, these results suggest that our drug repositioning approach can be used for identifying effective drug candidates, and *BUB1B, RRM2* and *ASF1B* are confirmed as druggable targets.

Discussion

ccRCC is a heterogeneous tumor that has been stratified into several molecular subtypes characterized by distinct mRNA expression patterns and opposite survival outcomes of patients.[14,41,42] Different subtypes of patients respond differently to chemotherapy. Low response rates and drug resistance exacerbate the challenges in ccRCC therapy. Thus, there is an urgent need for discovering new treatment options for patients who cannot benefit from the commonly used chemotherapies. In this study, we proposed an integrated approach combining the target prediction and drug repositioning for

DESeq2. *represents FDR<0.05. (b) Box plots showing the mRNA expression levels (TPM values) of favorable hub genes in TCGA tumor tissues compared to normal tissues. *represents FDR<0.05. (c) The log2 fold changes between the expression value of a hub gene in each TCGA tumor sample and the average expression value of this gene in TCGA normal samples. The x-axis represents individual tumor samples. (d) The IHC images of hub genes in normal tissues and tumor tissues. The images related to the same gene were conducted by the same antibody, which was described in the Method section. Scale bar, 50μm. (e) The essential scores of hub genes in 16 ccRCC cell lines. More negative scores indicate more essential for tumor cell survival and proliferation. (f) The mRNA expression levels (TPM values) of hub genes in the Caki-1 cell line. In the box plots, the bottom and top of the boxes represent the 25th and 75th percentiles. The whiskers represent the minimum and maximum values that are not outliers. The central band represents the median value.

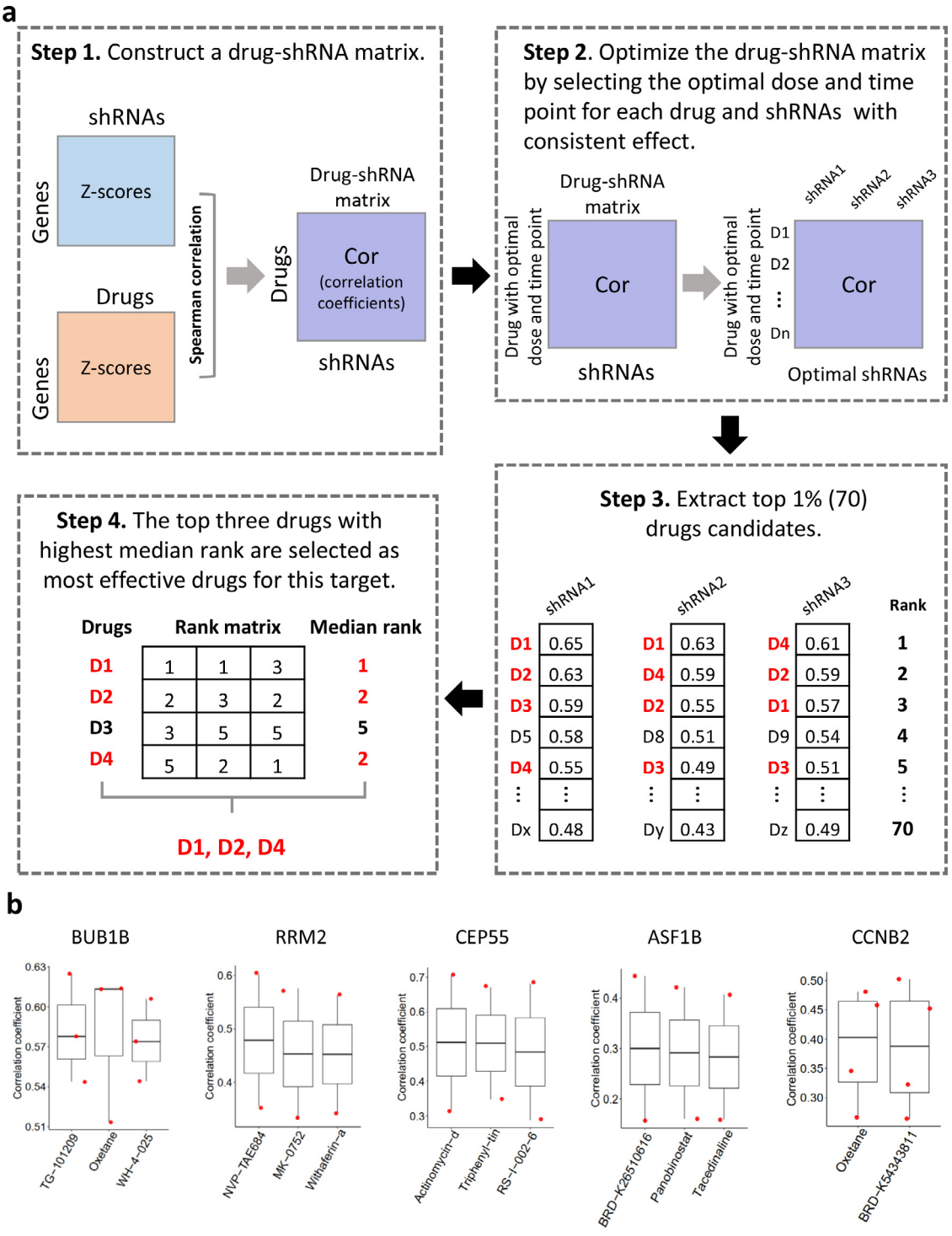

Figure 5. Drug repositioning for the target genes. (a) The flowchart of drug repositioning approach. The details have been described in Method section. (b) Box plots showing the three most effective drugs for each target gene. Each point represents the similarity (calculated as spearman correlation coefficient) between the repurposed drug perturbed effects and shRNA knockdown perturbed effects on cells. The bottom and top of the boxes represent the 25th and 75th percentiles. The whiskers represent the minimum and maximum values that are not outliers. The central band represents the median value.

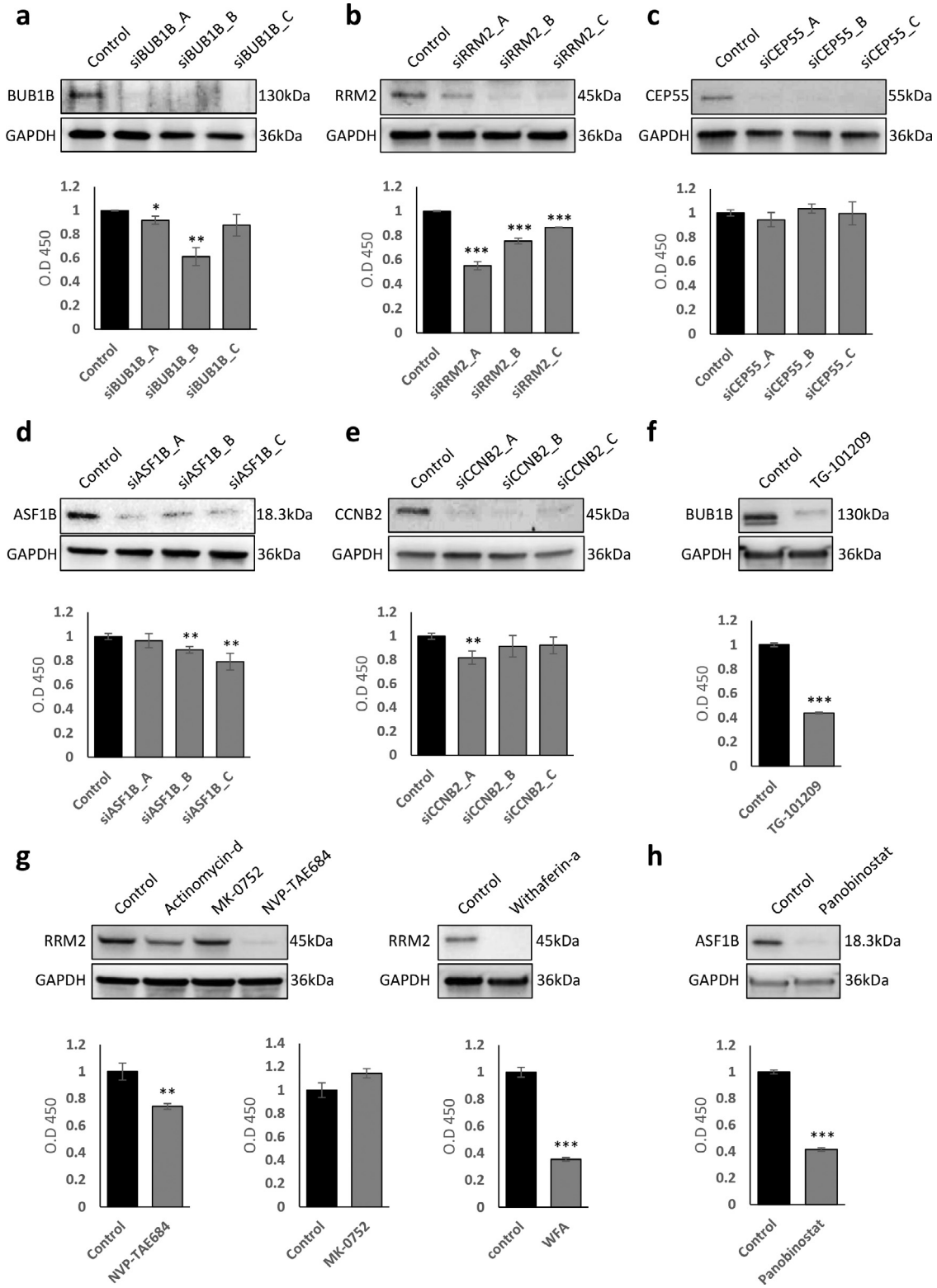

Figure 6. Validation of target genes and drug effects. (a–e) Western blots showing the protein levels of target genes in Caki-1 with siRNA transfection and negative control and bar plots showing the change of cell viability. (f–h) Western plots showing that the protein levels of target genes were inhibited by the treatment of their corresponding target drugs and bar plots showing the change of cell viability. The

ccRCC treatment. As a result, we identified four promising druggable target genes which encode the proteins involved in cell mitosis and cell cycle regulation, including *BUB1B, RRM2, ASF1B* and *CCNB2*, for treatment of ccRCC. *BUB1B* encodes a mitotic spindle checkpoint and exhibits a cell cycle dependent expression. It is undetectable in G1 and exhibits a gene expression peak in G2/M.[43,44] *RRM2* encodes ribonucleotide reductase M2 subunit and it is responsible for the ribonucleotide dezoxyribonucleotide conversion during the S phase of the cell cycle.[45] Thus, it is only expressed during the late G1 /early S phase and degraded in late S phase.[46] *CCNB2*, encoding cyclin B2, is involved in the G2-M transition in eukaryotes by activating CDC2 kinase[47–49] and it also shows a gene expression peak in G2/M.[44] *ASF1B*, encoding one of the isoforms of the histone H3−H4 chaperone anti-silencing function 1, is necessary for cell proliferation and differently expressed in the cycling and non-cycling cells.[50] It has been reported that *BUB1B* is an independent prognostic marker,[51] *RRM2* is a drug resistance-related marker,[52] *ASF1B* and *CCNB2* are metastasis markers for ccRCC.[53,54] These target genes are significantly upregulated in tumor tissues compared to the adjacent normal tissues in our analysis. Moreover, broad coverage of upregulated alteration for each of these target genes was observed in ccRCC patients (Figure 4c), implying that most patients may benefit from treatment of these target genes even disease states varied for individual patients.

Further, we developed a drug repositioning approach using the correlation analysis between shRNA-perturbed and drug-perturbed signatures and identified the three most effective drugs for each target. In the end, we validated the inhibitory effect of TG-101209 for targeting *BUB1B*, NVP-TAE684, and withaferin-a for targeting *RRM2*, and panobinostat for targeting *ASF1B* using *in vitro* model. TG-101209, a JAK2 inhibitor, was developed for patients with myeloproliferative disorders who carry the JAK2V617F mutation.[55] It also inhibits the tumor cell growth in myeloma[56] and lung cancer[57] using *in vitro* or *in vivo* models. NVP-TAE684, an ALK inhibitor, was designed to inhibit oncogenic ALK-rearranged fusion proteins (e.g., NPM-ALK).[58] It has been reported that NVP-TAE684 suppressed the cell proliferation in pancreatic adenocarcinoma[59] and neuroblastoma,[60] and reversed multidrug resistance in osteosarcoma.[61] Withaferin-a, a steroidal lactone derived from the medicinal plant *Withania somnifera* Dunal (Solanaceae), has a wide range of pharmacological activities, including cardioprotective, anti-inflammatory, and anti-angiogenesis.[62] It has been reported that withaferin-a induced cell apoptosis by inhibiting the expression or activation of STAT3 in ccRCC cells[63] and other cancer cells.[64–66] Panobinostat, a non-selective histone deacetylase inhibitor, is an FDA-approved oral drug to treat multiple myeloma.[67] The preclinical study demonstrated that panobinostat induced cell cycle arrest and apoptosis by mediating the dual degradation of Aurora A and B kinases in different renal cancer cell lines.[68] The drug effect of panobinostat is worthwhile to further explore for early-stage ccRCC patients by single-drug therapy or combination therapy even through a phase 2 trial, including only 20 advanced ccRCC patients showed no benefit from panobinostat.[69] Since all the three druggable targets, *BUB1B, RRM2* and *ASF1B*, are cell cycle dependent genes, we speculate that the combination of their corresponding repurposed drugs could generate a synergistic effect on the inhibition of the tumor cell growth. On the other hand, the synergistic effect by combining our repurposed drugs with the clinically used first-line chemotherapy drugs (e.g., tyrosine kinase inhibitors, mTOR inhibitors and monoclonal antibodies against VEGF) is also worthwhile to investigate in the future work. Interestingly, the treatment effects of the combination of panobinostat with tyrosine inhibitor axitinib[70] or proteasome inhibitor bortezomib[71] have been demonstrated *in vitro* or *in vivo* ccRCC models.

The integrated approach we proposed in this study could be broadly applied to other cancers. There are several requirements for the application of this approach. First, at least two independent RNA-seq datasets should be guaranteed to extract the repeatable disease-specific signature genes. The reliability of gene expression data determines the accuracy of the target genes we predicted. In this study, we collected the two RNA-seq datasets from two ccRCC cohorts, TCGA and Japanese cohorts. At the beginning, we observed quite different numbers of prognostic genes (7,813 and 1,335 genes) in these two cohorts. One reason is that the Japanese cohort only includes 100 samples whereas the TCGA cohort includes 528 samples, which leads to relatively low statistical power during survival analysis in Japanese cohort. More importantly, since the patients in the two cohorts have very different races, geographical characteristics and culture backgrounds, these heterogeneities also contribute to the difference. Even under this situation, the two sets of prognostic genes showed significantly high concordance. Moreover, we found the common functional modules in the two cohorts based on co-expression network analysis. These evidences suggested that there is a common tumor-driven molecular mechanism controlled by a set of hub genes in ccRCC. Secondly, the target-specific shRNA-perturbed signature

cells in the negative control group were only treated by DMSO. The change of cell viability between the two groups was compared by two-sided T-test. Four replicates were set in each group. Error bar represents standard deviation. *represents $P < 0.05$. **represents $P < 0.01$, ***represents $P < 0.001$.

profiles are necessary for our drug repositioning analysis. In this study, we extracted the shRNA-perturbed signatures from the CMap data portal. Researchers could use their own shRNA-perturbed gene expression profiles to generate the signatures as well. In addition, overexpression-perturbed data is an excellent alternative to replace the shRNA-perturbed data. Notably, a drug is considered to inhibit a target gene if these target-specific overexpression-perturbed signatures are negatively correlated with drug-perturbed signatures. In conclusion, we proposed a useful approach for disease-target prediction and drug-target prediction. We also successfully validated the essentiality of target genes and drug efficacy using an *in vitro* model. This study provides new insight into the treatment of ccRCC and promotes precision medicine.

Contributors

X.L. contributed to the conceptualization, formal analysis, methodology, visualization and writing-original draft in this study. K.S. contributed to the methodology, visualization and validation. W.K., M.Y. and H.Y. contributed to the visualization and validation. Y.S., H.K. and S.O. contributed to the data curation and resources. H.T. contributed to the validation and writing-review&editing. S.S. contributed to the validation, software and writing-review&editing. J.B. contributed to the validation and software. J.N. and M.U. contributed to the resources and software. C.Z. contributed to the supervision, conceptualization and writing-review&editing. A. M contributed to the supervision, project administration, funding acquisition and writing-review&editing. All the authors read and approved the final manuscript.

Data sharing statement

The RNA-seq data we used in this study were downloaded from public data resources. We have clarified all the accession codes in the Method section. The codes used in this study were available in https://github.com/sysmedicine.

Declaration of interests

X.L. works as a consultant for Bash Biotech Inc., San Diego, CA, USA. H.T., A.M., and S.S. are the founders and shareholders of Bash Biotech Inc., San Diego, CA, USA. Bash Biotech Inc. filed a patent application about the use of TG-101209, NVP-TAE684, withaferin-a and panobinostat in the treatment of renal cell carcinoma. The other authors declare no competing interests.

Acknowledgements

This research was funded by Knut and Alice Wallenberg Foundation grant number CJDB 72110 and Bash Biotech Inc., San Diego, CA, USA. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2019-3-599.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.103963.

References

1 Motzer RJ, Jonasch E, Boyle S, et al. NCCN guidelines insights: kidney cancer, version 1.2021. *J Natl Compr Cancer Netw*. 2020;18(9):1160–1170.
2 National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology. 2020; published online July 15. https://www.nccn.org/professionals/physician_gls/pdf/kidney.pdf. Accessed 1 July 2021.
3 Hsieh JJ, Purdue MP, Signoretti S, et al. Renal cell carcinoma. *Nat Rev Dis Primers*. 2017;3:17009.
4 Luo H, Li M, Yang M, Wu FX, Li Y, Wang J. Biomedical data and computational models for drug repositioning: a comprehensive review. *Brief Bioinform*. 2021;22(2):1604–1619.
5 Turanli B, Altay O, Boren J, et al. Systems biology based drug repositioning for development of cancer therapy. *Semin Cancer Biol*. 2021;68:47–58.
6 Altay O, Mohammadi E, Lam S, et al. Current status of COVID-19 therapies and drug repositioning applications. *iScience*. 2020;23:(7) 101303.
7 Nagaraj AB, Wang QQ, Joseph P, et al. Using a novel computational drug-repositioning approach (DrugPredict) to rapidly identify potent drug candidates for cancer treatment. *Oncogene*. 2018;37(3):403–414.
8 Xuan P, Cao Y, Zhang T, Wang X, Pan S, Shen T. Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics*. 2019;35(20):4108–4119.
9 Yang M, Wu G, Zhao Q, Li Y, Wang J. Computational drug repositioning based on multi-similarities bilinear matrix factorization. *Brief Bioinform*. 2021;22(4):bbaa267.
10 Zerbini LF, Bhasin MK, de Vasconcellos JF, et al. Computational repositioning and preclinical validation of pentamidine for renal cell cancer. *Mol Cancer Ther*. 2014;13(7):1929–1941.
11 Koudijs KKM, Terwisscha van Scheltinga AGT, Bohringer S, Schimmel KJM, Guchelaar HJ. Personalised drug repositioning for clear cell renal cell carcinoma using gene expression. *Sci Rep*. 2018;8(1):5250.
12 Lamb J, Crawford ED, Peck D, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;313(5795):1929–1935.
13 Stathias V, Turner J, Koleti A, et al. LINCS data portal 2.0: next generation access point for perturbation-response signatures. *Nucleic Acids Res*. 2020;48(D1):D431–D4D9.
14 Li X, Kim W, Juszczak K, et al. Stratification of patients with clear cell renal cell carcinoma to facilitate drug repositioning. *iScience*. 2021;24:(7) 102722.
15 Pornputtapong N, Nookaew I, Nielsen J. Human metabolic atlas: an online resource for human metabolism. *Database (Oxf)*. 2015;2015:bav068.
16 Mardinoglu A, Boren J, Smith U, Uhlen M, Nielsen J. Systems biology in hepatology: approaches and applications. *Nat Rev Gastroenterol Hepatol*. 2018;15(6):365–377.
17 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–674.
18 Tatlow PJ, Piccolo SR. A cloud-based workflow to quantify transcript-expression levels in public cancer compendia. *Sci Rep*. 2016;6:39259.
19 Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525–527.
20 Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44(8):e71.
21 Sato Y, Yoshizato T, Shiraishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet*. 2013;45(8):860–867.

22 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–842.
23 Meyers RM, Bryan JG, McFarland JM, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49(12):1779–1784.
24 Ghandi M, Huang FW, Jane-Valbuena J, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature*. 2019;569(7757):503–508.
25 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
26 Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287.
27 Pons P, Latapy M. Computing Communities in Large Networks Using Random Walks. In: *Proceedings of the Computer and Information Sciences-ISCIS*. Berlin, Heidelberg: Springer; 2005.
28 Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, complex systems*. 2006;1695(5):1–9.
29 Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437–1452. e17.
30 Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347:(6220) 1260419.
31 Uhlen M, Bjorling E, Agaton C, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteom*. 2005;4(12):1920–1932.
32 Uhlen M, Zhang C, Lee S, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357(6352).
33 Li X, Kim W, Arif M, et al. Discovery of functional alternatively spliced PKM transcripts in human cancers. *Cancers (Basel)*. 2021;13(2).
34 Han JD. Understanding biological functions through molecular networks. *Cell Res*. 2008;18(2):224–237.
35 Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. *Science*. 2003;302(5643):249–255.
36 Wu J, Xia X, Hu Y, Fang X, Orsulic S. Identification of infertility-associated topologically important genes using weighted co-expression network analysis. *Front Genet*. 2021;12: 580190.
37 Bidkhori G, Benfeitas R, Klevstig M, et al. Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes. *Proc Natl Acad Sci USA*. 2018;115(50):E11874–E11E83.
38 Palmer D, Fabris F, Doherty A, Freitas AA, de Magalhaes JP. Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. *Aging (Albany NY)*. 2021;13 (3):3313–3341.
39 Bozhilova LV, Pardo-Diaz J, Reinert G, Deane CM. COGENT: evaluating the consistency of gene co-expression networks. *Bioinformatics*. 2021;37(13):1928–1929.
40 Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ. Multidimensional drug profiling by automated microscopy. *Science*. 2004;306(5699):1194–1198.
41 The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013;499(7456):43–49.
42 Brannon AR, Reddy A, Seiler M, et al. Molecular stratification of clear cell renal cell carcinoma by consensus clustering reveals distinct subtypes and survival patterns. *Genes Cancer*. 2010;1(2):152–163.
43 Davenport JW, Fernandes ER, Harris LD, Neale GA, Goorha R. The mouse mitotic checkpoint gene bub1b, a novel bub1 family member, is expressed in a cell cycle-dependent manner. *Genomics*. 1999;55(1):113–117.
44 Whitfield ML, Sherlock G, Saldanha AJ, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell*. 2002;13(6):1977–2000.
45 Grolmusz VK, Karaszi K, Micsik T, et al. Cell cycle dependent RRM2 may serve as proliferation marker and pharmaceutical target in adrenocortical cancer. *Am J Cancer Res*. 2016;6(9):2041–2053.
46 Zhang K, Hu S, Wu J, et al. Overexpression of RRM2 decreases thrombspondin-1 and increases VEGF production in human cancer cells *in vitro* and *in vivo*: implication of RRM2 in angiogenesis. *Mol Cancer*. 2009;8:11.
47 Draetta G, Luca F, Westendorf J, Brizuela L, Ruderman J, Beach D. Cdc2 protein kinase is complexed with both cyclin A and B: evidence for proteolytic inactivation of MPF. *Cell*. 1989;56(5):829–838.
48 Petri ET, Errico A, Escobedo L, Hunt T, Basavappa R. The crystal structure of human cyclin B. *Cell Cycle*. 2007;6(11):1342–1349.
49 Wu T, Zhang X, Huang X, Yang Y, Hua X. Regulation of cyclin B2 expression and cell cycle G2/m transition by menin. *J Biol Chem*. 2010;285(24):18291–18300.
50 Corpet A, De Koning L, Toedling J, et al. Asf1b, the necessary Asf1 isoform for proliferation, is predictive of outcome in breast cancer. *EMBO J*. 2011;30(3):480–493.
51 Sekino Y, Han X, Kobayashi G, et al. BUB1B overexpression is an independent prognostic marker and associated with CD44, p53, and PD-L1 in renal cell carcinoma. *Oncology*. 2021;99(4):240–250.
52 Osako Y, Yoshino H, Sakaguchi T, et al. Potential tumorsuppressive role of microRNA99a3p in sunitinibresistant renal cell carcinoma cells through the regulation of RRM2. *Int J Oncol*. 2019;54 (5):1759–1770.
53 Jiangqiao Z, Tao Q, Zhongbao C, et al. Anti-silencing function 1B histone chaperone promotes cell proliferation and migration via activation of the AKT pathway in clear cell renal cell carcinoma. *Biochem Biophys Res Commun*. 2019;511(1):165–172.
54 Peng R, Wang Y, Mao L, Fang F, Guan H. Identification of core genes involved in the metastasis of clear cell renal cell carcinoma. *Cancer Manag Res*. 2020;12:13437–13449.
55 Pardanani A, Hood J, Lasho T, et al. TG101209, a small molecule JAK2-selective kinase inhibitor potently inhibits myeloproliferative disorder-associated JAK2V617F and MPLW515L/K mutations. *Leukemia*. 2007;21(8):1658–1668.
56 Ramakrishnan V, Kimlinger T, Haug J, et al. TG101209, a novel JAK2 inhibitor, has significant *in vitro* activity in multiple myeloma and displays preferential cytotoxicity for CD45+ myeloma cells. *Am J Hematol*. 2010;85(9):675–686.
57 Sun Y, Moretti L, Giacalone NJ, et al. Inhibition of JAK2 signaling by TG101209 enhances radiotherapy in lung cancer models. *J Thorac Oncol*. 2011;6(4):699–706.
58 Galkin AV, Melnick JS, Kim S, et al. Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. *Proc Natl Acad Sci USA*. 2007;104(1):270–275.
59 Duong HQ, Than VT, Nguyen HT, et al. Anaplastic lymphoma kinase inhibitor NVP-TAE684 suppresses the proliferation of human pancreatic adenocarcinoma cells. *Oncol Rep*. 2021;45(4):1–11.
60 Schonherr C, Ruuth K, Kamaraj S, et al. Anaplastic Lymphoma Kinase (ALK) regulates initiation of transcription of MYCN in neuroblastoma cells. *Oncogene*. 2012;31(50):5193–5200.
61 Ye S, Zhang J, Shen J, et al. NVP-TAE684 reverses multidrug resistance (MDR) in human osteosarcoma by inhibiting P-glycoprotein (PGP1) function. *Br J Pharmacol*. 2016;173(3):613–626.
62 Mohan R, Hammers HJ, Bargagna-Mohan P, et al. Withaferin A is a potent inhibitor of angiogenesis. *Angiogenesis*. 2004;7(2):115–122.
63 Um HJ, Min KJ, Kim DE, Kwon TK. Withaferin A inhibits JAK/STAT3 signaling and induces apoptosis of human renal carcinoma Caki cells. *Biochem Biophys Res Commun*. 2012;427(1):24–29.
64 Choi BY, Kim BW. Withaferin-A inhibits colon cancer cell growth by blocking STAT3 transcriptional activity. *J Cancer Prev*. 2015;20 (3):185–192.
65 Lee J, Hahm ER, Singh SV. Withaferin A inhibits activation of signal transducer and activator of transcription 3 in human breast cancer cells. *Carcinogenesis*. 2010;31(11):1991–1998.
66 Yco LP, Mocz G, Opoku-Ansah J, Bachmann AS. Withaferin A inhibits STAT3 and induces tumor cell death in neuroblastoma and multiple myeloma. *Biochem Insights*. 2014;7:1–13.
67 Laubach JP, Moreau P, San-Miguel JF, Richardson PG. Panobinostat for the treatment of multiple myeloma. *Clin Cancer Res*. 2015;21(21):4767–4773.
68 Cha TL, Chuang MJ, Wu ST, et al. Dual degradation of aurora A and B kinases by the histone deacetylase inhibitor LBH589 induces G2-M arrest and apoptosis of renal cancer cells. *Clin Cancer Res*. 2009;15(3):840–850.
69 Hainsworth JD, Infante JR, Spigel DR, Arrowsmith ER, Boccia RV, Burris HA. A phase II trial of panobinostat, a histone deacetylase inhibitor, in the treatment of patients with refractory metastatic renal cell carcinoma. *Cancer Investig*. 2011;29(7):451–455.
70 Rausch M, Weiss A, Zoetemelk M, et al. Optimized Combination of HDACI and TKI efficiently inhibits metabolic activity in renal cell carcinoma and overcomes sunitinib resistance. *Cancers (Basel)*. 2020;12(11).
71 Sato A, Asano T, Isono M, Ito K, Asano T. Panobinostat synergizes with bortezomib to induce endoplasmic reticulum stress and ubiquitinated protein accumulation in renal cancer cells. *BMC Urol*. 2014;14:71.