



ARTICLE

Identification of important genes and drug repurposing based on clinical-centered analysis across human cancers

Ying Li¹, Ya-ping Dong^{1,2}, You-wen Qian³, Le-xing Yu¹, Wen Wen^{1,4}, Xiu-liang Cui^{1,4} and Hong-yang Wang^{1,4,5}

Identification of the functional impact of mutated and altered genes in cancer is critical for implementing precision oncology and drug repurposing. In recent years, the emergence of multiomics data from large, well-characterized patient cohorts has provided us with an unprecedented opportunity to address this problem. In this study, we investigated survival-associated genes across 26 cancer types and found that these genes tended to be hub genes and had higher K-core values in biological networks. Moreover, the genes associated with adverse outcomes were mainly enriched in pathways related to genetic information processing and cellular processes, while the genes with favorable outcomes were enriched in metabolism and immune regulation pathways. We proposed using the number of survival-related neighbors to assess the impact of mutations. In addition, by integrating other databases including the Human Protein Atlas and the DrugBank database, we predicted novel targets and anticancer drugs using the drug repurposing strategy. Our results illustrated the significance of multidimensional analysis of clinical data in important gene identification and drug development.

Keywords: clinical correlation; survival outcomes; somatic mutation; drug repurposing; network analysis; pancancer.

Acta Pharmacologica Sinica (2021) 42:282–289; <https://doi.org/10.1038/s41401-020-0451-1>

INTRODUCTION

High-throughput genomic technologies have greatly facilitated cancer research, and recent consortium projects such as The Cancer Genome Atlas (TCGA) have provided a preliminary understanding of the landscape of genetic alterations in cancers [1]. A fundamental challenge in developing novel cancer therapies is the identification of driver genes from a large pool of candidate genes with distinct mutations or expression patterns [2]. Large-scale loss-of-function screens based on RNAi or CRISPR-Cas9 techniques have been conducted to explore cancer driver genes [3–5]. Recently, Ng et al. developed a moderate-throughput functional genomic platform and annotated more than 1000 cancer variants of unknown significance [6].

Several computational algorithms have been developed to distinguish driver genes from nonessential passenger variants. Traditional population-based strategies were designed to look for genes that have more mutations than expected compared with the average background mutation rate at various levels [7]. Considering that different mutations of a certain gene may exert diverse functions, a variety of algorithms were proposed to interpret the detailed function of a single mutation at the subgene level [8]. For instance, several tools were developed to detect mutation hotspots based on three-dimensional protein structures [9–11]. Both experimental and computational methods are powerful tools for assessing the importance of certain genes or mutations. However, they are not sufficient for the investigation of clinical significance.

In oncology practice, clinical characteristics reflect the profiles of cancer patients and serve as important guides for cancer diagnosis, classification and treatment. Clinical characteristic-related genes and mutations are more likely to be drivers and targets for cancer treatment. Moreover, the development of cancer can only rarely be attributed to an individual molecule. Instead, most clinical characteristics arise from complex interactions between the cell's numerous constituents. Therefore, it is necessary to analyze molecules related to the clinical characteristic at the systematic level and to integrate multidimensional data such as mRNA expression, copy number variation (CNV) and mutation data. Drug repurposing, the application of an existing therapeutic to a new disease indication, has become a promising and cost-effective strategy for cancer drug development [12, 13]. For example, metformin, which is currently the first-line drug treatment for type 2 diabetes has exhibited a potential role in multiple cancers [14]. Survival-related genes provide targets for drug repurposing.

In this study, we performed an association analysis between patient survival outcomes (OS) and integrated multidimensional data including RNA sequencing data and copy number profiles. Survival-related genes were identified in 26 cancer types from the TCGA project [1]. Then, we outlined the functional and network properties of these genes. We assessed the mutation influence based on the number of clinical characteristics-related neighbors. Moreover, by integrating data from The Human Protein Atlas (THPA) [15] and DrugBank [16], we repurposed several drugs for anticancer treatment.

¹National Center for Liver Cancer, Second Military Medical University, Shanghai 200438, China; ²The Graduate School of Fujian Medical University, Fuzhou 350122, China;

³Department of Pathology, Eastern Hepatobiliary Surgery Hospital/Institute, Second Military Medical University, Shanghai 200433, China; ⁴International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Hospital/Institute, Second Military Medical University, Shanghai 200433, China and ⁵Ministry of Education (MOE) Key Laboratory on Signaling Regulation and Targeting Therapy of Liver Cancer, Second Military Medical University, Shanghai 200433, China

Correspondence: Xiu-liang Cui (wafyai@163.com) or Hong-yang Wang (hywangk@vip.sina.com)

These authors contributed equally: Ying Li and Ya-ping Dong

Received: 31 December 2019 Accepted: 30 May 2020

Published online: 18 June 2020

MATERIALS AND METHODS

Tumor and normal samples and data sets

Twenty-six major cancer types were chosen for our analysis. Detailed information is provided in Supplementary Table S1. The high-throughput data obtained by next-generation sequencing, including mRNA expression, CNV and gene mutation data, and the clinical information for cancer patients were downloaded from the TCGA public access web portal (<https://cancergenome.nih.gov/>). The details of the TCGA sequencing platforms and data processing are as follows. Level 3 values normalized by the Expectation Maximization (RSEM) algorithm were employed as the expression levels of the corresponding mRNAs. Somatic mutation data were generated by Illumina GA/HiSeq automated DNA sequencing platform and SOLiD sequencing platform. For CNV, the segmented data were provided by the Broad Institute Affymetrix Genome-Wide Human SNP Array 6.0. GISTIC 2.0 was employed to calculate the gene-based CNV value [17] and analyze the raw TCGA data with amplification and deletion cutoffs of 0.3 on a log base 2 scale. We used the in-house script to map multidimensional data and clinical information. Cancer-related genes, including oncogenes and tumor suppression genes, were obtained from the COSMIC database.

Drug target data

Information on drugs and their targets was obtained from the DrugBank database (Version 5) [16], and the xml file was parsed using an in-house script. There are 2021 FDA-approved drugs and over 6000 experimental drugs and 4661 targets in the database. The first level of ATC classification codes, which indicate drug therapeutic properties was used to label the identified drugs.

Protein expression and protein-protein interaction network

Protein expression level data in normal and tumor tissues were obtained from THPA [15]. Protein expression was classified into 4 levels: high, median, low and undetected. The distribution of cancer patients at each level was calculated, and the expression level was defined as that in the maximum proportion of tumor patients. The Human Protein Reference Database (HPRD) (release 9) is a widely used protein-protein interaction (PPI) database that contains 9270 proteins and 39,241 interactions [18] and Rolland et al. provided a systematic map of 13,944 high-quality binary interactions between 4303 proteins [19]. We integrated the two resources to construct the background PPI network in this study, and there were 49,584 interactions between 11,041 proteins in our integrated PPI network.

Pathway enrichment analysis

Pathway enrichment analysis was performed using the R 'clusterProfiler' package (v. 3.6.0) [20]. Genes with $P < 0.05$ were considered significantly related to prognosis, and in cancer types that contained more than 1000 significantly related genes, the top 1000 most correlated genes were used for pathway enrichment analysis.

Statistical analysis

To assess the association of each dimension of genes with survival outcomes, we first classified patients into different groups and compared the survival between these groups. For mRNA expression analysis, patients were equally divided into two groups according to their expression level. For DNA mutation analysis, patients were classified into mutated and nonmutated groups. For CNV analysis, patients were classified into amplification, normal and deletion groups. Patient survival between subgroups was assessed using the log-rank test with the R 'survival' package (v. 2.37). Ten-fold cross-validation was used to demonstrate the significance of the identified survival-related genes. In more detail, 90% of patients from the whole cohort were randomly selected, and the association between expression and patient survival was

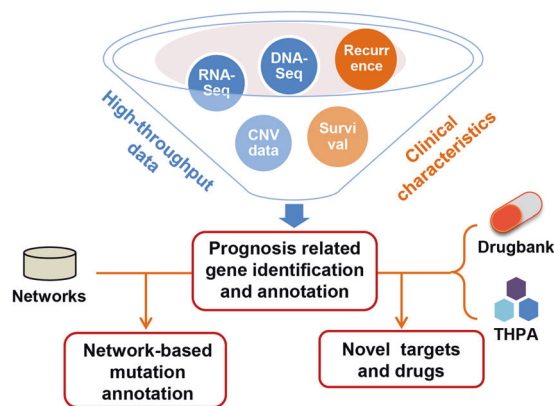


Fig. 1 Scheme of data analysis. The prognosis-related genes were identified by calculating the correlation between multiple types of molecular data (gene expression and CNV) and patient survival and recurrence in 26 cancer types. Then, by integrating data from other resources such as the DrugBank (drug data) and the THPA (protein expression data) databases, we identified driver mutations and novel targets and drugs.

calculated using the method described above. The procedure was repeated 100 times. As shown in Supplementary Fig. S1, the result was robust according to the cohort size. The FDR value was calculated using R software to perform multiple test correction. All the source code of our analysis is available under GNU Public License v3 through GitHub at <https://github.com/vsfarmer/4CR>.

RESULTS

Overview of data analysis

Data from 26 major cancer types with sufficient sample sizes (more than 50 samples) from TCGA were used for our analysis. For each cancer type, we only reserved samples with 2 types of molecular data (copy number variations [CNVs] and mRNA expression) together with clinical information, and there were nearly 9000 cases in our analysis (detailed information in Supplementary Table S1). Genes for which the expression, mutation and CNV data were correlated with patient OS were identified. Then we characterized the prognosis-related genes using pathway annotation and network analysis. The number of prognosis-related neighbors was used to assess the impact of frequently mutated genes on patient prognosis. By integrating data from other resources, such as drug target and protein expression data from the THPA database, we showed the advantages of clinical-centered analysis for target identification and drug development, as shown in Fig. 1.

Identification and annotation of survival-related genes

We first investigated the survival rates for patients with 26 cancer types in TCGA. As shown in Fig. 2a, prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA) and thymoma (THYM) exhibited better prognosis, while acute myeloid leukemia (AML), glioblastoma multiforme (GBM) and mesothelioma (MESO) displayed worse prognosis according to the average 1-year survival rate. However, as the survival rates were derived from the TCGA database, they did not reflect the disease prevalence. Then, we calculated the correlation between mRNA expression and CNV levels and patient OS (detailed information is provided in the Materials and methods). Genes for which high expression or amplification was associated with better survival were referred to as favorable outcome genes, and those for which high expression was associated with worse survival were referred to as adverse outcome genes. To explore the function of survival-

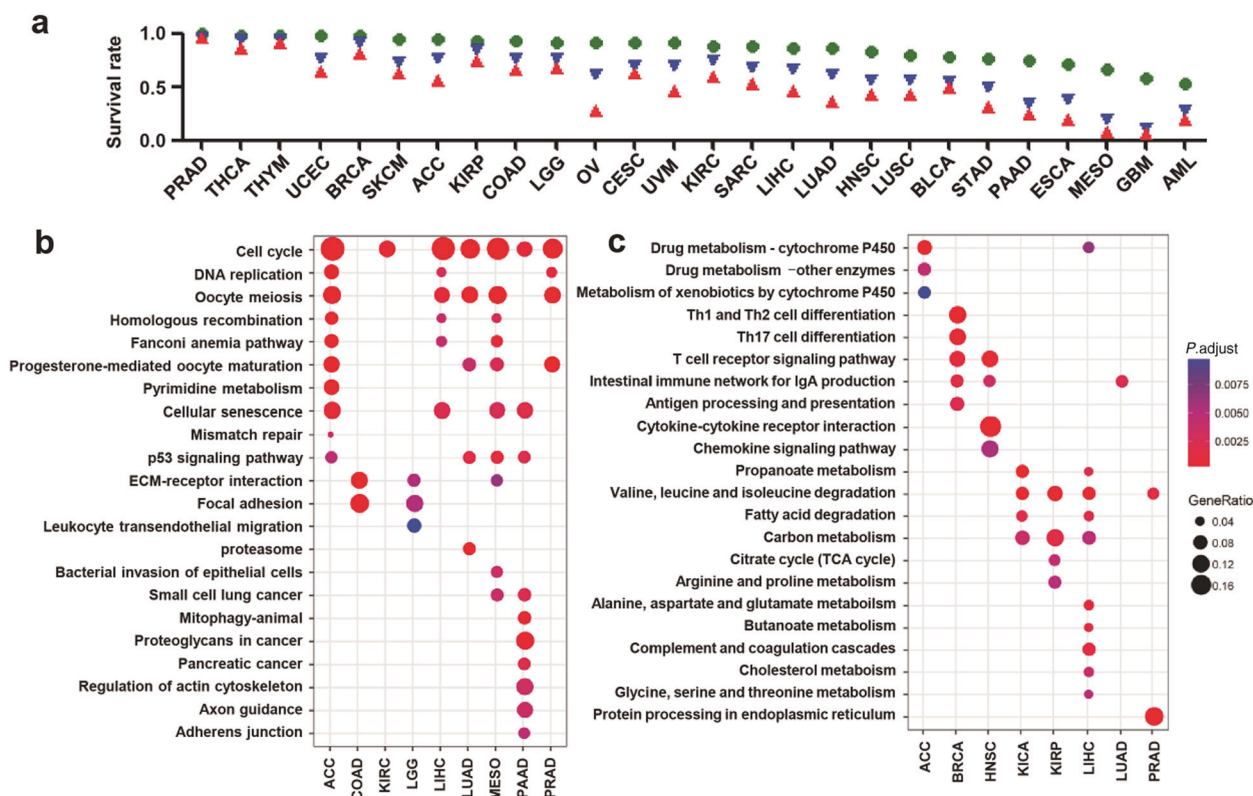


Fig. 2 Analysis of prognosis-related genes. **a** 1-year survival rate (green line), 3-year survival rate (blue line) and 5-year survival rate (red line) of patients with 26 cancer types in TCGA. **b** Pathways enriched by genes with adverse outcomes. **c** Pathways enriched by genes with favorable outcomes.

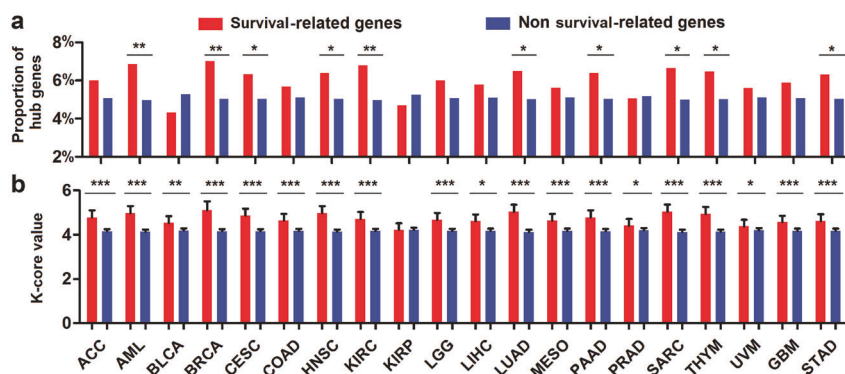


Fig. 3 Network topology of survival-related genes. **a** Hub proportion of survival-related genes (red bar) and non-survival-related genes (blue bar). **b** K-core value of survival-related genes (red bar) and non-survival-related genes (blue bar).

related genes, we performed pathway enrichment analysis for the genes for which the expression was related to survival. As shown in Fig. 2b, it is not surprising that the adverse outcome genes were mainly enriched in genetic information processing and cellular process pathways, such as the cell cycle and DNA replication, as the activation of these pathways can induce the proliferation of tumor cells. In contrast, the favorable outcome genes were mainly enriched in immune pathways, such as the T cell receptor signaling pathway, which indicated that the activation of the immune system can improve the OS of cancer patients. In addition, metabolic pathways, such as valine, leucine and isoleucine degradation, were also enriched as shown in Fig. 2c.

Functionally important genes are usually topologically important nodes, such as hubs, in biological networks [21]. However, it

was reported that in 4 cancer types prognostic genes are unlikely to be hubs in the coexpression network [22]. Here, we investigated the network topological properties of genes for which the mRNA expression levels were correlated with OS in the PPI. Six cancer types were excluded due to their limited number of survival-related genes (less than 10 prognostic genes in each cancer type). The connectivity distributions were examined and the top 5% node genes with the highest connectivity were defined as hub genes, according to the literature [22]. We found that the proportion of hub genes among the survival-related genes was higher than that among the non survival-related genes in 17 out of 20 cancer types and the difference was significant in 10 cancer types, as shown in Fig. 3a. The K-core is a concept in graph theory, and the K-core of a graph G is a maximal subgraph of G in which all vertices have a degree of at least K. It has been

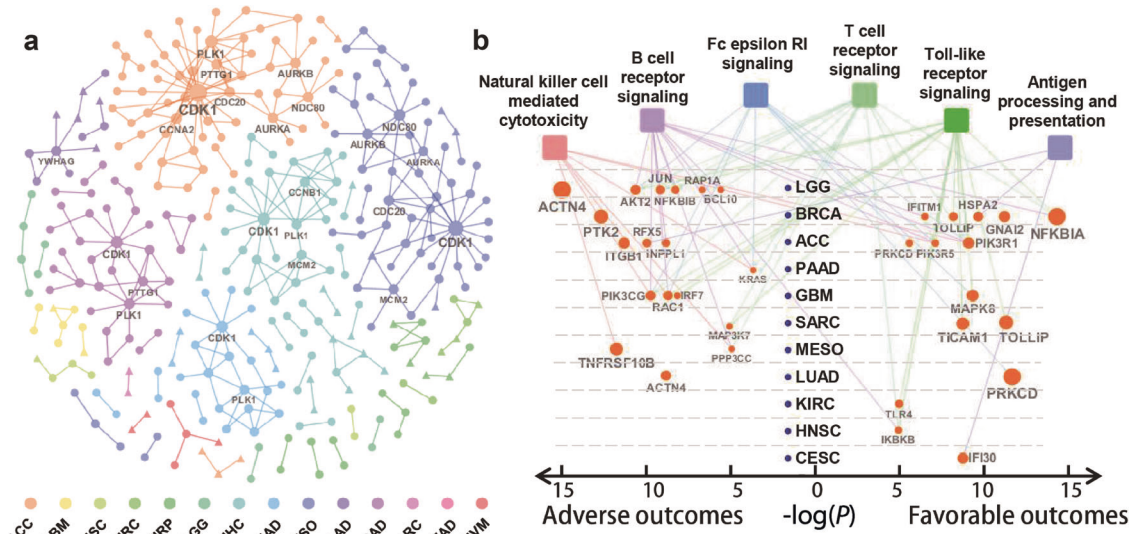


Fig. 4 Functional annotation of survival-related genes. **a** Interaction network of survival-related genes. Circular nodes represent genes with favorable outcomes, while "V" nodes represent genes with adverse outcomes; node colors represent cancer types. **b** Interaction network of survival-related genes and immune pathways. The gene (orange circle) and its corresponding pathway (rectangle node) were linked.

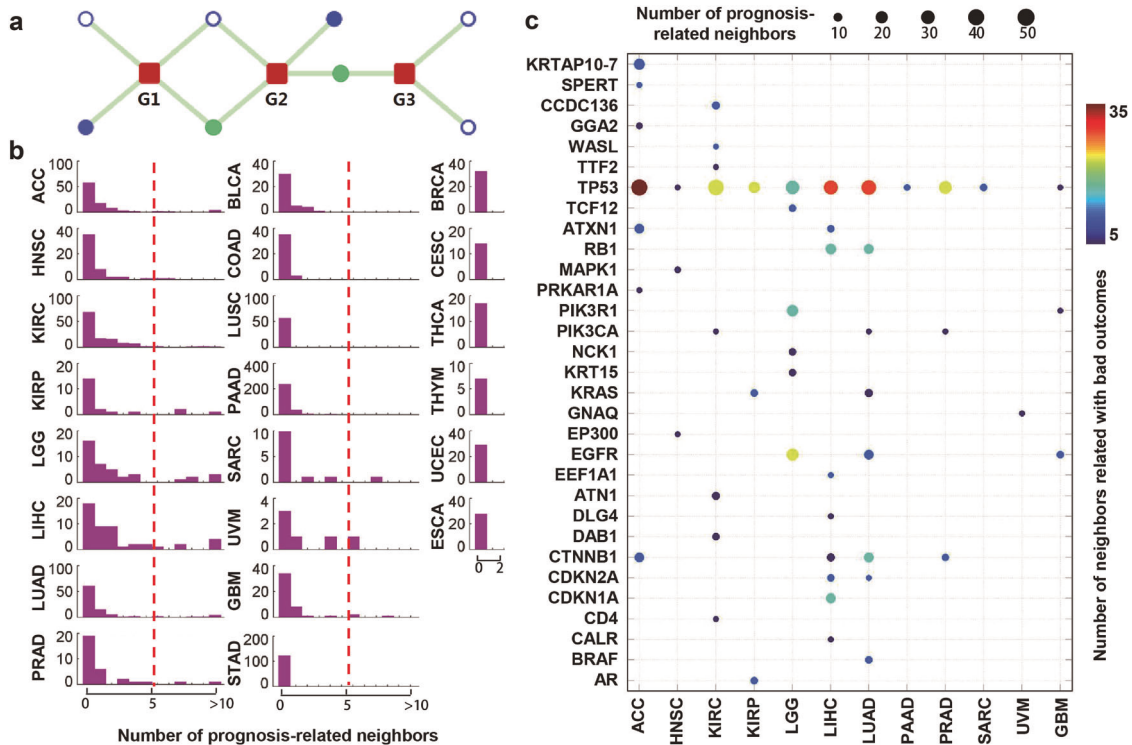


Fig. 5 Assessment of mutation influence based on network neighbor analysis. **a** The illustration of N_{exp}^{surv} . The red rectangular nodes (G1, G2 and G3) represent the frequently mutated genes and the circular nodes represent network neighbors. The green solid nodes represent neighbors associated with favorable outcomes, the blue solid nodes represent neighbors associated with bad outcomes and the hollow circle nodes represent non-prognosis-related neighbors. **b** Distribution of N_{exp}^{surv} number of frequently mutated genes in 22 cancer types. **c** Bubble diagram of mutated genes with $N_{exp}^{surv} \geq 5$. The size of the nodes represents the N_{exp}^{surv} of mutated genes, and the color of nodes represents the N_{exp}^{bad} of mutated genes in the corresponding cancer type.

applied to evaluate the coreness of a gene in biological networks. We calculated the K-core values of survival-related genes and non-survival-related genes. As shown in Fig. 3b, the prognostic genes had higher K-core values than other genes in all the 20 cancer types, and the difference was significant in 19 cancer types.

The interaction network of the genes correlated with patient survival is shown in Fig. 4a ($P < 0.01$, FDR < 0.1 , detailed information in Supplementary Table S2). As expected, well-known cancer-related genes including *CDK1*, *PLK1*, *CCNA2*, *AURKA* and *AURKB* were related to the poor outcomes in multiple cancer types and these genes occupied key positions in the network.

Immune pathways play important roles in cancer development and immunotherapy has become a promising strategy for cancer treatment [23]. In our analysis, six immune pathways were obtained from KEGG, including natural killer cell-mediated cytotoxicity, B cell receptor signaling, Fc epsilon RI signaling, T cell receptor signaling, Toll-like receptor signaling and antigen processing and presentation pathways. The survival-related genes enriched in immune pathways are shown in Fig. 4b, and these genes may serve as potential biomarkers for immunotherapy [24]. Prognostic genes of other categories including nuclear receptors, cytokine receptors, enzyme-linked receptors, CD molecules and cell adhesion molecules (CAMs) and their ligands were also investigated (Supplementary Fig. S2). In addition, we found that the majority of the prognosis-related genes encoded proteins that are categorized as cytokine receptors and enzyme-linked receptors or receptor tyrosine kinases (RTKs),

such as *ROR1*, *EPHB4*, *PTK7*, *AXL* and *TIE1*. In addition, CD molecules serve as cell surface makers of immune cells and play important roles in the recognition, adhesion and activation of T cells and B cells. According to our analysis, 17 CD molecules were identified to be correlated with cancer OS, among which 13 molecules were associated with adverse outcomes. Moreover, 13 CAM genes (*ITGB1*, *ITGB8*, *ADAM12*, *MIA*, *FN1*, *HSPG2*, *ICAM3* and *MADCAM1*) were prognosis-related, 8 of which belonged to the integrin family; the other 4 genes (*ROBO4*, *LRFN3*, *DLG4* and *PTPRB*) belonged to the immunoglobulin superfamily.

Functional assessment of mutations based on survival-related network neighbors
We hypothesized that the mutated genes with more survival-related network neighbors are more likely to be clinically influential than those with fewer survival-related network

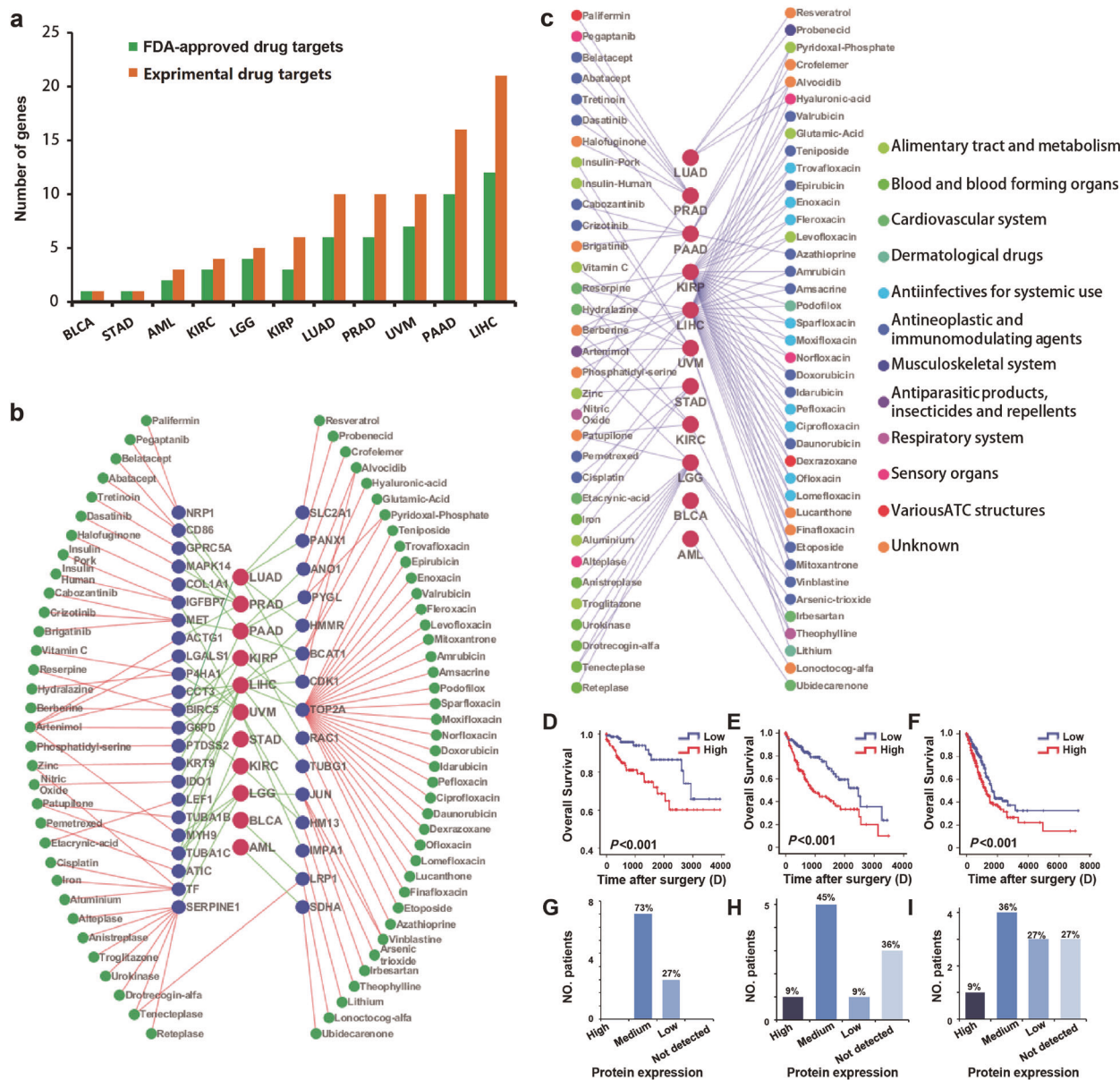


Fig. 6 Drug repurposing based on prognosis-related target analysis. **a** The distribution of AORP genes targeted by FDA-approved drugs and experimental compounds according to the DrugBank database in different cancer types. **b** The mapping between FDA-approved drugs (green nodes) and their clinically actionable genes (blue nodes), and the correlation between adverse outcome (red line) and recurrence (blue line) of these clinically actionable genes across different cancer types (red nodes). **c** Interaction network of different cancers and repurposed drugs. Overall survival rates of KIRP (**d**), LIHC (**e**) and LUAD (**f**) patients with high or low expression of BIRC5. Protein expression in KIRP (**g**), LIHC (**h**) and LUAD (**i**) tumor tissues.

neighbors. Therefore, we proposed a survival-related neighbor-based method to assess the importance of mutations. For a given mutated gene, the number of OS-related neighbors (N_{exp}^{surv}) and the number of neighbors of poor outcome-related neighbors (N_{exp}^{bad}) were analyzed. As illustrated in Fig. 5a, the N_{exp}^{surv} values of G1, G2 and G3 were 2, 3 and 1 respectively, and the N_{exp}^{bad} values of G1, G2 and G3 were 1, 2 and 1, respectively. Thus, using MutSigCV [25], we first identified the frequently mutated genes for each cancer type and then calculated their N_{exp}^{surv} scores. The distribution of N_{exp}^{surv} scores among the highly mutated genes is shown in Fig. 5b. Moreover, we screened mutated genes with $N_{exp}^{surv} \geq 5$ for each cancer type and 51 unique mutated genes were identified. As shown in Fig. 5c, the *TP53* mutation had a high N_{exp}^{surv} score in 9 cancer types including adrenocortical carcinoma (ACC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), brain lower-grade glioma (LGG), liver hepatocellular carcinoma (LIHC), LUAD, PRAD, pancreatic adenocarcinoma (PAAD) and sarcoma (SARC). Our results indicated that the mutation of *TP53* was common in diverse types of human cancer and also affected the expression of multiple downstream survival-related genes [26]. In addition to *TP53*, other well-known oncogenes and tumor suppressor genes, such as *PIK3CA*, *CTNNB1* and *EGFR*, also had high N_{exp}^{surv} scores in multiple cancer types. Furthermore, we found that the N_{exp}^{surv} scores of cancer-related genes from the COSMIC database were higher than those of other genes ($P < 1 \times 10^{-11}$, Wilcoxon rank-sum test). Our results were consistent with the publicly accepted fact that these genes may be determining factors for patient OS. It is noteworthy that in addition to the well-studied mutations, there are novel mutations predicted to have a large impact on patient OS, and these merit further investigation. Tumor mutation burden (TMB) was recently recognized as a biomarker for the efficacy of PD-1 treatment in lung cancer [27, 28]. We investigated the relationship between TMB and 7 CPCs, as shown in Supplementary Fig. S3. TMB was related to patient age, TNM stage and smoking. Interestingly, patients with a history of other malignancies were prone to having a higher TMB. However, in our study, we found that although relapsed patients had a higher TMB, there was no significant correlation between TMB and overall survival. Our findings were consistent with previous reports that patients whose tumors were ultramutated displayed better overall survival in several cancer types, such as glioma [29] and endometrial cancer [30]. This is probably because the immune system was activated by the antigen.

Drug repurposing based on survival-related target analysis

It is crucial to identify the appropriate targets for anticancer treatment. We detected candidate drug targets using the following criteria: (i) genes that were associated with patient adverse outcomes (AOGs); (ii) genes for which the protein expression levels were higher in tumors than normal tissues; and (iii) genes that were known targets of approved drugs in clinical use. Following the identification of AOGs for each cancer type, we investigated their protein expression levels in tumor and normal tissues using the THPA database. AOGs whose protein expression levels were higher in tumor tissues than in the corresponding normal tissues were selected. Drugs targeting these genes were screened from the DrugBank database [16]. The distribution of these genes targeted by FDA-approved drugs and experimental compounds in different cancer types is shown in Fig. 6a. Among them, 38 genes were targeted by 73 FDA-approved drugs, and the identified genes mainly encoded enzymes, cell adhesion molecules (CAMs) and channel proteins. The interaction network between FDA-approved drugs, target genes and corresponding cancer types is shown in Fig. 6b. Furthermore, we built direct connections between drugs and cancer types based on their common target genes, as shown in Fig. 6c. The first level of Anatomical Therapeutic Chemical (ATC)

classification codes, which indicate drug therapeutic properties was used to label the identified drugs. As shown in Fig. 6d, 21 drugs were intended for cancer treatment (ATC classification code was L: antineoplastic and immunomodulating agents), such as crizotinib, dasatinib and belatacept. The majority of predicted drugs (51 drugs) were not intended for cancer treatment but may be repurposed for cancer therapy in the near future according to our analysis. In fact, the anticancer activities of certain drugs identified in our analysis were consistent with previous reports. For example, Campos-Arroyo et al. reported that probenecid could sensitize neuroblastoma cells, including tumor cells with stem features, to the effects of cisplatin [31]. Jones et al. found that a metastatic colorectal cancer patient exhibited a profound and durable response after treatment with irbesartan, indicating the possibility of the repurposing of irbesartan as an anticancer therapy [32]. Moreover, hydralazine has been shown to reverse doxorubicin resistance in breast cancer [33]. These reports further confirmed our findings, suggesting the potential value of our analysis for novel cancer drug development based on drug repositioning strategies.

These results indicate the potential of our analysis for identifying appropriate cancer targets and provide helpful clues for drug repurposing. For instance, baculoviral IAP repeat containing 5 (*BIRC5*) was associated with patient survival. As shown in Fig. 6d–f, the KIRP, LIHC and LUAD patients in the *BIRC5* high group exhibited shorter survival times than those patients in the *BIRC5* low group. Moreover, *BIRC5* protein expression was not detected in normal renal, liver and lung tissues according to the data from THPA database, and the *BIRC5* expression level was higher in the corresponding tumor tissues than in the normal tissues, as shown in Fig. 6g–i. We also observed that the mRNA expression in tumors was significantly higher than that in normal tissues (Supplementary Fig. S4). *BIRC5* encodes survivin, which is a small inhibitor of apoptosis (IAP) that regulates the senescence, migration, and invasion of cancer cells. Survivin is considered a promising target in multiple cancers [34]. According to the DrugBank database, there are two approved drugs targeting *BIRC5*, berberine and reserpine. Berberine has been used orally for various parasitic and fungal infections and as an antidiarrheal medicine, and reserpine has been used as an antihypertensive and antipsychotic agent. Our data indicate that they may be repurposed for the treatment of these cancers.

DISCUSSION

The identification of determinant genes is a critical challenge for precision oncology [2–6, 13], and correlations between genes and clinical characteristics could serve as important indicators for identifying driver genes. In the past few years, the accumulation of multidimensional sequencing data from large, well-characterized cancer cohorts has provided us with an unprecedented opportunity to address these issues [1]. Recently, the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium has provided a landscape of genome alteration of 38 cancers [35, 36]. In this study, we performed a comprehensive, pancancer analysis to investigate the correlation between multidimensional molecular data and patient OS across cancer types.

Specifically, we focused on survival-related genes that are involved in the immune signaling pathways. Our analysis suggested that the activation of these pathways could improve patient OS. Previous studies have identified survival-related cytokines and chemokines that showed significant pan-cancer prognostic ability based on mRNA and CNV separately [37, 38]. Herein, we used the two omics datasets simultaneously, and the candidate genes tended to be cancer type specific. It is generally considered that functionally important genes are also topologically important in biological networks [21]. Consistently, we found

that survival-related genes were prone to be hub genes and displayed higher K-core values than the non-survival-related genes. These results could deepen our understanding of the function of these genes in cancer development.

Interestingly, our data showed that although somatic mutations accumulated with age and were associated with TNM staging and history of malignancy as well as environmental factors such as tobacco smoking, there was no correlation between somatic mutation number and overall survival of cancer patients, which indicates the complexity of the correlation between cancer mutations and patient OS [29, 30]. Furthermore, we introduced a network-based index N_{exp}^{surv} to assess the impact of frequently mutated genes on patient OS, and identified unique mutated genes that may exert determinantal function on patient OS in multiple cancer types. As expected, the well-known oncogenes and tumor suppressor genes had higher N_{exp}^{surv} scores than other genes.

Association analysis between survival-related genes and drug targets is a widely used strategy for drug repositioning. In this study, we assumed that the AOGs could be used as candidate targets, and screened drugs targeting these genes. It is worth noting that drugs not intended for cancer treatment could also be repurposed for cancer therapy according to our analysis, which provides helpful clues for further studies. For example, we found that ubidecarenone targets SDHA, which is AOG of AML. Ubidecarenone is a powerful antioxidant and a lipid-soluble and essential cofactor in mitochondrial oxidative phosphorylation that is currently used as a dietary supplement [39]. Our data indicate that ubidecarenone could be promising for AML treatment. Drugs targeting BIRC5 were also predicted to be effective in KIRP, LIHC and LUAD. However, the limitation of our study was that we did not validate our findings.

Collectively, we conducted systematic clinical-centered analyses of mutation impact and drug repurposing. The genes with adverse outcomes and the genes with favorable outcomes were enriched in different pathways, and the survival-related genes tended to occupy important positions in biological networks. Genomic mutations are considered the driving causes of cancers. We proposed a network-based index N_{exp}^{surv} to assess the functional importance of mutations. Our analysis not only identified influential mutations but also provided possible mechanisms. By integrating the drug target and protein expression data together with OS correlation results, we identified therapeutic targets, and our approach promotes drug development based on the drug repurposing strategy.

ACKNOWLEDGEMENTS

This study was funded by the National Natural Science Foundation of China (81602620, 81502086, 81572896, 8172203 and 91859205). National Major Scientific and Technological Special Project for "Significant New Drugs Development" (2018ZX09101-002), the Scientific Research Foundation of Shanghai Municipal Commission of Health and Family Planning (No. 20154Y0140) and the Shanghai Pujiang Program (18PJD060).

AUTHOR CONTRIBUTIONS

YL and YPD performed the centrality analysis, drafted and revised the manuscript. YWQ, LX Y and WW participated in the statistical analysis and manuscript writing. XLC and HYW conceived and supervised this study.

Compliance with ethical standards

ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41401-020-0451-1>) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

REFERENCES

- Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA. Cancer Genome Atlas Research Network et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet.* 2013;45:1113–20.
- Chen JC, Alvarez MJ, Talos F, Dhruv H, Rieckhof GE, Iyer A, et al. Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell.* 2014;159:402–14.
- McDonald ER III, de Weck A, Schlabach MR, Billy E, Mavrakis KJ, Hoffman GR, et al. Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. *Cell.* 2017;170:577–92 e10.
- Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. *Cell.* 2017;170:564–76 e16.
- Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang CZ, Ben-David U, et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* 2016;6:914–29.
- Ng PK, Li J, Jeong KJ, Shao S, Chen H, Tsang YH, et al. Systematic functional annotation of somatic mutations in cancer. *Cancer Cell.* 2018;33:450–62 e10.
- Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet.* 2015;47:106–14.
- Porta-Pardo E, Kamburov A, Tamborero D, Pons T, Grases D, Valencia A, et al. Comparison of algorithms for the detection of cancer drivers at subgene resolution. *Nat Methods.* 2017;14:782.
- Gao JJ, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med.* 2017;9:4.
- Niu B, Scott AD, Sengupta S, Bailey MH, Batra P, Ning J, et al. Protein-structure-guided discovery of functional mutations across 19 cancer types. *Nat Genet.* 2016;48:827–37.
- Tokheim C, Bhattacharya R, Niknafs N, Gygyax DM, Kim R, Ryan M, et al. Exome-scale discovery of hotspot mutation regions in human cancer using 3D protein structure. *Cancer Res.* 2016;76:3719–31.
- Jin GX, Wong STC. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov Today.* 2014;19:637–44.
- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov.* 2004;3:673–83.
- Ben Sahra I, Le Marchand-Brustel Y, Tanti JF, Bost F. Metformin in cancer therapy: a new perspective for an old antidiabetic drug? *Mol Cancer Ther.* 2010;9:1092–9.
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science.* 2015;347:1260419.
- Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 2014;42:D1091–7.
- Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukheim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12:R41.
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database–2009 update. *Nucleic Acids Res.* 2009;37:D767–72.
- Rolland T, Tasan M, Charlotiaux B, Pevzner SJ, Zhong Q, Sahni N, et al. A proteome-scale map of the human interactome network. *Cell.* 2014;159:1212–26.
- Yu GC, Wang LG, Han YY, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic.* 2012;16:284–7.
- Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;5:101–13.
- Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat Commun.* 2014;5:3231.
- Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160:48–61.
- Bullock TNJ. TNF-receptor superfamily agonists as molecular adjuvants for cancer vaccines. *Curr Opin Immunol.* 2017;47:70–7.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499:214–8.
- Hollstein M, Sidransky D, Vogelstein B, Harris CC. p53 mutations in human cancers. *Science.* 1991;253:49–53.
- Hellmann MD, Callahan MK, Awad MM, Calvo E, Ascierto PA, Atmaca A, et al. Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer. *Cancer Cell.* 2018;33:853–61 e4.
- Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med.* 2018;378:2093–104.

29. Erson-Omay EZ, Caglayan AO, Schultz N, Weinhold N, Omay SB, Ozduman K, et al. Somatic POLE mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis. *Neuro Oncol.* 2015;17:1356–64.
30. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. *Nature.* 2013;497:67–73.
31. Campos-Arroyo D, Maldonado V, Bahena I, Quintanar V, Patino N, Carlos Martinez-Lazcano J, et al. Probenecid sensitizes neuroblastoma cancer stem cells to cisplatin. *Cancer Invest.* 2016;34:155–66.
32. Jones MR, Schrader KA, Shen Y, Pleasance E, Ch'ng C, Dar N, et al. Response to angiotensin blockade with irbesartan in a patient with metastatic colorectal cancer. *Ann Oncol.* 2016;27:801–6.
33. Segura-Pacheco B, Perez-Cardenas E, Taja-Chayeb L, Chavez-Blanco A, Revilla-Vazquez A, Benitez-Bribiesca L, et al. Global DNA hypermethylation-associated cancer chemotherapy resistance and its reversion with the demethylating agent hydralazine. *J Transl Med.* 2006;4:32.
34. Altieri DC. Survivin, cancer networks and pathway-directed drug discovery. *Nat Rev Cancer.* 2008;8:61–70.
35. Consortium I T P-C A o W G. Pan-cancer analysis of whole genomes. *Nature.* 2020;578:82–93.
36. Mitra R, Adams CM, Jiang W, Greenawalt E, Eischen CM. Pan-cancer analysis reveals cooperativity of both strands of microRNA that regulate tumorigenesis and patient survival. *Nat Commun.* 2020;11:968.
37. Wong HS, Chang CM, Liu X, Huang WC, Chang WC. Characterization of cytokinome landscape for clinical responses in human cancers. *Oncoimmunology.* 2016;5:e1214789.
38. Wong HS, Chang WC. Losses of cytokines and chemokines are common genetic features of human cancers: the somatic copy number alterations are correlated with patient prognoses and therapeutic resistance. *Oncoimmunology.* 2018;7:e1468951.
39. Carli F, Chiellini EE, Bellich B, Macchiavelli S, Cadelli G. Ubidecarenone nanoemulsified composite systems. *Int J Pharm.* 2005;291:113–8.