

Development of Prognostic Signature Based on Pan-cancer Proteomics

Weiguo Huang

Wenzhou Medical University First Affiliated Hospital

Wanqing Weng

Wenzhou Medical University First Affiliated Hospital

Yukai Xiang

Wenzhou Medical University First Affiliated Hospital

Hongqi Shi (✉ shhoqi@163.com)

Wenzhou Medical University First Affiliated Hospital <https://orcid.org/0000-0002-8740-6585>

Primary research

Keywords: proteomics, pan-cancer, biomarker

Posted Date: May 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-30886/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Utilizing genomic data to predict cancer prognosis was insufficient. Proteomics can improve our understanding of the etiology and progression of cancer and improve the assessment of cancer prognosis. Based on CPTAC (Clinical Proteomic Tumor Analysis Consortium) which has generated extensive proteomics data of the vast majority of tumors, we can perform a proteomic pan-carcinoma analysis.

Methods: The proteomics data and clinical features of cancer patients were collected from CPTAC. We screened 69 differentially expressed proteins with R software. GO and KEGG analysis were performed to clarify the function of these proteins. The DEPs-based prognostic model was identified by least absolute shrinkage and selection operator (LASSO)-Cox regression model. The time-dependent receiver operating characteristics analysis was used to evaluate the ability of the prognostic model to predict overall survival.

Results: A total of 69 differentially expressed proteins were screened in five different types of cancers: hepatocellular carcinoma (HCC), lung adenocarcinoma (LUAD), children's brain tumor tissue consortium (CBTTC), clear cell renal cell carcinoma (CCRC) and uterine corpus endometrial carcinoma (UCEC). Furthermore, the differentially expressed proteins were related to cell metabolism, cell proliferation and extracellular matrix. Then 24 DEPs-based classifiers for predicting OS was developed by LASSO-Cox regression model in training cohort, which was validated in validation cohort.

Conclusions: In the present study, we identified DEPs-based survival-predictor model to predict most cancers. We are the first group to utilize proteomics to construct a pan-cancer prognosis model, which could accurately and effectively predict the survival rate of most cancers.

Introduction

As the most prevalent fatal disease, cancer ranked second in all mortality worldwide in 2017.¹ And the death rate of cancer was increasing year by year, cancer deaths increased from 7.62 million in 2007 to 9.56 million in 2017. In 2018, 18.1 million people worldwide have been diagnosed with various types of cancer.² Despite the significant progress in treatment, timely diagnosis and high cost of treatment make it impossible to obtain effective treatment, which was still the reason for the low 5-year survival rate of most cancers.³ In order to develop optimal anti-cancer treatment protocols and elucidate the mechanism of tumorigenesis, it is essential to estimate the prognosis of tumor patients.⁴ Although many studies used RNA sequence data from the Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) to evidence many tumor prognostic biomarkers and construct many prognostic models,^{5,6} utilizing genomic data to predict cancer prognosis was insufficient and imprecise.

It is widely acknowledged that tumor cells were characterized by rapid generation and abnormal proliferation. Hence, tumor tissues would regulate the expression of proteins and promote the production

of proteins associated with cancer progression.⁷ Moreover, proteins were the functional effectors of cellular processes as well as the targets for a vast majority of therapeutics.⁸ Therefore, the study of proteomics can improve our understanding of cancer aetiology and progression as well as heighten the assessment of cancer prognosis.⁹ Although most previous studies have focused on the effects of individual specific protein on cancer prognosis,¹⁰⁻¹² cancer is a heterogeneity disease that does not only involve individual protein but also interactions among proteins of different function. The Clinical Proteomic Tumor Analysis Consortium (CPTAC) project had generated a great deal of proteomics data of the vast majority of tumors by mass spectrometry.¹³ Based on the proteomics data from CPTAC, we expect to combine multiple proteins to construct a pan-cancer prognostic model.

In current study, we screened out differentially expressed proteins (DEPs) in five cancers: hepatocellular carcinoma (HCC), uterine corpus endometrial carcinoma (UCEC), children's brain tumor tissue consortium (CBTTC), lung adenocarcinoma (LUAD) and clear cell renal cell carcinoma (CCRC). Next, we explored the role of the differentially expressed proteins in cancer and the relationships among them. Furthermore, the DEPs-based survival-predictor model was also developed for predicting survival rates for the vast majority of cancers.

Methods

Patient datasets

The proteomic data of HCC, CBTTC, CCRC, LUAD and UCEC were extracted from The CPTAC (<https://proteomics.cancer.gov/programs/cptac>) in November 2019.

Identification of DEPs between tumor tissues and adjacent nontumorous tissues

For the proteomic data from CPTAC, background correction, quantile normalization and batch normalization were performed using R software (version 3.6.1). The protein expression values of these five cancers were normalized by the “sva” package. The bioconductor (<http://www.bioconductor.org>) package “limma” was employed for DEP screening. A $|\log_2\text{Fold Change}| > 1$ and an adjusted P value < 0.05 were set as cut-off criteria.

Functional enrichment analyses

We performed KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis and Gene ontology (GO) analysis using R package “enrichplot”, “enrichplot”, “GOplot”.

PPI network construction

The PPI network of DEPs was performed by STRING (version 11.0) and a combined score > 0.9 (high confidence) was set as the cut-off criterion. Using cytoscape online software (<http://www.cytoscape.org/>) to visualize the results from STRING.

Construction of DEPs-based classifiers

Based on univariate Cox regression models, we identify single DEP as independent prognostic DEPs for OS with p value < 0.05 . The least absolute shrinkage and selection operator (LASSO)-Cox regression model¹⁴ was used to identify the most accurate predictive DEPs for OS. The correlation of each prognostic DEPs was performed by R package “ggcorrplot”, “statn”.

Predictive performance of the DEPs-based classifiers.

The patient's risk score is obtained by multiplying the expression of DEPs in LASSO by their respective coefficients. And the patients were stratified into two risk-groups by median. The survival was analyzed by the Kaplan–Meier log rank analysis. The time-dependent receiver operating characteristics (tdROC) analysis was used to assess performance of single DEP and classifiers through the “timeROC” package of R software. The area under the curve (AUC) of tdROC reflected predictive accuracy. P -values < 0.05 were considered statistically significant.

Data analysis

The Student's t test, Wilcoxon test, and other data processing were completed by SPSS 19.0. Kaplan-Meier analysis is calculated by the “survminer” package of R software. When all the hypotheses are $P < 0.05$, the difference is statistically significant.

Flow chart

The workflow was shown in Figure 1.

Results

Differentially expressed proteins in five Cancers

Firstly, we acquired five types of cancer of proteomic data sets from the CPTAC data portal, which contained HCC, CBTTTC, CCRC, LUAD and UCEC. According to the criteria of $\log_2 | \text{FoldChange} | > 1$ and $\text{FDR} < 0.05$, we identified 69 differentially expressed proteins (DEPs) between tumor tissues and adjacent nontumorous tissues using “limma”, and then plotted volcano and heat maps (Figure 2A, B). Among the 69 proteins, 26 proteins expression were upregulated in cancerous tissues such as Cyclin-dependent kina (CDK1) and Proliferation marker protein Ki-67 (MKI67), while 43 proteins were down-regulated in cancerous tissues such as Beta-enolase (ENO3) and Glycerol-3-phosphate dehydrogenase [NAD(+)] (GPD1) (Figure 2B).

GO analysis and KEGG analysis

In order to explore the role of the 69 DEPs in tumors, we conducted GO analysis and KEGG analysis. And the 69 DEPs were mainly associated with the following biological processes: carboxylic acid biosynthetic

process, organic acid biosynthetic process, G1/S transition of mitotic cell cycle, cell cycle G1/S phase transition, monocarboxylic acid biosynthetic process, glucose metabolic process, hexose metabolic process, and DNA replication (Figure 3A). The results also indicated that the DEPs were mainly associated with the following cellular contents: nuclear chromosome part, extracellular matrix, telomeric region and MCM complex (Figure 3A). Besides, the DEPs were related to molecular functions, such as extracellular matrix structural constituent, carbohydrate binding, helicase activity and monosaccharide binding (Figure 3A). Similar to GO analysis, KEGG analysis showed the DEPs primarily contributed to the following pathways: Cell cycle, Glycolysis / Gluconeogenesis, DNA replication, Carbon metabolism, Pentose phosphate pathway and Fructose and mannose metabolism (Figure 3B). Furthermore, combining GO cluster diagram and GO chord diagram, we found that the parts of DEPs involved in DNA replication, Cell cycle and Arginine and proline metabolism were mainly high-expressed, and others associated with these GO terms such as Carbon metabolism and Fructose and mannose metabolism were both highly and poorly expressed (Figure 3C, D).

DEPs Interaction Clusters Common across Five Cancers

The 69 DEPs were used for the network analysis and almost half the DEPs formed an interaction network after eliminating proteins that acted independently (Figure 4A). And these interacting proteins were roughly separated into four groups with CDK1, ENO3, Argininosuccinate synthase (ASS1) and Versican core protein (VCAN) as the cores (Figure 4A, B). CDK1 was observed to be the key hub protein that interacted with DNA replication licensing factor MCM2 (MCM2), DNA replication licensing factor MCM3 (MCM3), DNA replication licensing factor MCM4 (MCM4), DNA replication licensing factor MCM5 (MCM5), DNA replication licensing factor MCM6 (MCM6), DNA replication licensing factor MCM7 (MCM7), MKI67, Ribonucleoside-diphosphate reductase subunit M2 (RRM2), TRIP13, 14-3-3 protein sigma (SFN), Histone H1.5 (HIST1H1B), cAMP-dependent protein kinase type II-beta (PRKAR2B). ENO3 interacted with Hexokinase-2 (HK2), Fructose-1,6-bisphosphatase isozyme 1 (FBP1), Fructose-1,6-bisphosphatase isozyme 2 (FBP2), Fructose-bisphosphatase aldolase B (ALDOB) and Phosphoglucomutase-like protein 5 (PGM5). VCAN interacted with Aspartoacylase (ASPA), Decorin (DCN), Thrombospondin-2 (THBS2), Tenascin-X (TNXB), Lymphatic vessel endothelial hyaluronic acid receptor 1 (LYVE1) and Mimecan (OGN). ASS1 interacted with PGM5, ASPA, Alpha-aminoadipate aminotransferase (AADAT), pyrroline-5-carboxylate reductase 1 (PYCR1), and Alpha-aminoadipic semialdehyde synthase (AASS).

The effect of individual DEPs on Survival

To explore the effect of these proteins on cancer prognosis, Kaplan-Meier survival analyses were performed using individual protein. Based on the median value of each DEP expression, we divided the cancer patients into two clusters: high protein level and low protein level. Then we defined four types of cancer as the training cohort: HCC, CCRC, LUAD and UCEC; and defined CBTTTC as validation cohort. We counted the OS of patients from the training cohort. As shown in Supplementary Figure 1, only ten proteins out of 69 DEPs were statistically significant in the survival analysis ($P < 0.05$). Patients whose

cancerous tissue expressed higher levels of one of RRM2, Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 (PLOD2), MKI67, MCM5, and CKD1 had lower survival rates (Supplementary Figure 1A-E). And Patients whose cancerous tissue expressed higher levels of one of FBP1, FBP2, ENO3, GPD1, and ASS1 had higher survival rates (Supplementary Figure 1F-J). Yet regrettably, receiver operating characteristic (ROC) analysis of them were unsatisfactory: RRM2 (1 year AUC=0.622), PLOD2 (1 year AUC=0.635), MKI67 (1 year AUC=0.617), MCM5 (1 year AUC=0.595), CKD1 (1 year AUC=0.610), FBP1 (1 year AUC=0.323), FBP2 (1 year AUC=0.320), ENO3 (1 year AUC=0.383), GPD1 (1 year AUC=0.362), ASS1 (1 year AUC=0.437) (Supplementary Figure 2). Although 3 year AUC of PLOD2 reached 0.722, 1 year and 2 year AUC were unsatisfactory. In summary, although the ten proteins can be used as biomarkers of cancer prognosis, none of them could accurately predict OS.

DEPs-based survival-predictor model constructing

For acquiring a more excellent model, multiple DEPs were combined to predict survival rates for cancer patients. We first conducted univariate Cox analyses in training cohort and found that 33 DEPs related to survival were identified (Figure 5A). Then we used sixty-nine DEPs to perform the LASSO Cox regression model in training cohort. Based on the results of the LASSO Cox regression model, 24 prognostic DEPs with non-zero regression coefficients were finally chosen as the potential prognostic biomarkers for the OS of cancer patients (Figure 5B, C). The detailed information of DEPs for constructing the prognostic signature was summarized in Table 1. The formula of the twenty-four-DEPs survival-predictor model was as follows: twenty-four-DEPs survival-predictor model score = (0.303235530256179*MKI67)+(0.259559228152558*LOXL2)+(0.216349150569518*PLIN4)+(0.163857099694478*IL33)+(0.153385186100743*MDK)+(0.144674735753098*P4HA2)+(0.13190528953757*AKR7A3)+(0.121054348420759*PLCXD3)+(0.120550067398402*CDK1)+(0.077423785033028*SRPX)+(0.0692670634423047*PRPH)+(0.0633036451678804*PRKAR2B)+(0.0468066907473914*P4HA1)+(0.0467283261834873*CALML3)+(0.0342288464237997*SFN)+(0.00251963795312595*DES)-(0.021873780076824*PHYHD1)-(0.041717955104614*GPD1)-(0.0516581125556701*AADAT)-(0.0740355402044938*PGM5)-(0.165116242778278*ADH1C)-(0.245540744438086*FBP2)-(0.265391369780318*ENO3)-(0.388025693935519*EHD3). The correlation of each protein in the 24-DEPs model was shown in Figure 5D and 5E. Among these proteins, the values of correlation between CDK1 and MKI67, P4HA2 and P4HA1, PGM5 and IL33, PGM5 and DES were all more than 0.5.

Evaluation of the survival-predictor model

Based on the survival-predictor model, we evenly divided cancer patients into two groups by the median risk score cut-off point, which value is 0.250379: High risk and Low risk (Figure 6A). The patient information was shown in Table 2 and Table 3. Furthermore, the expression heatmap of the 24 DEPs in high risk or low risk group was shown in Figure 6A. We then estimated the accuracy of the 24-DEPs model on predicting survival. The Kaplan-Meier survival curves showed that survival rates were significantly lower in the High risk (P <0.001) (Figure 6B). The ROC analysis showed the one, two, and three years AUC of the 24-DEPs survival-predictor model were 0.764, 0.754, and 0.742 respectively (Figure

6C). Remarkably, the AUC of the 24-DEPs survival-predictor model was more than the AUC of the ten proteins described above (Supplementary Figure 2). So, compared with a single protein as a predictor, the 24-DEPs survival prediction model had accurate and powerful prediction capability.

In order to further validate the availability of this model, we used the same 24-DEPs survival-predictor model and cut-off point to cluster patients in validation cohort (CBTTC) (Figure 6D). And the survival analysis also indicated that high risk had a worse OS ($P < 0.001$) (Figure 6E). The result of the ROC analysis was also satisfactory: one year AUC=0.724, two years AUC=0.689, three years AUC=0.671 (Figure 6F). In conclusion, the 24 DEPs-based classifiers could accurately predict the survival not only in the training cohort, but also in the validation cohort.

Discussion

As a complex disease, cancer involves not only in DNA alterations, but also in protein expression and modification.¹⁵ With technological improvements, CPTAC generates comprehensive mass spectrometry-based proteomic data for most cancers,¹³ which providing a unique opportunity for pan-cancerous proteomics research with sufficient data.

In current study, we firstly screened 69 differentially expressed proteins in five types of cancer tissue. More importantly, the expression trend of the DEPs was consistent in all five cancers, which indicated these proteins were not specific to any certain type of cancer. Among the DEPs, CDK1 played an important role in progression into mitotic phase, which could drive the cell cycle in all cell types.¹⁶ Previous studies also showed that CDK1 expression was upregulated in a majority of tumor tissues, which correlated with the prognosis of cancer patients.¹⁷⁻¹⁹ And MCM2, MCM3, MCM4, MCM5, MCM6, MCM7 formed the MiniChromosome Maintenance 2-7 complex, which was exported by the CDKs to trigger DNA replication.²⁰ In brief, CDK1 interacted with MCM2-7 complex to participate in the cell cycle, which was the same as the GO analysis and KEGG analysis. Furthermore, we found CDK1, as a key hub protein, interacted with other DEPs to form an interaction cluster. In addition to MCM2-7 complex, other proteins in the cluster also influenced the growth and division of tumor cells by participating in the cell cycle such as RRM2, PRKAR2B, and MKI67.²¹⁻²³ Most DEPs related to the cell cycle were up-regulated, which was consistent with the vigorous growth and division of tumor cells. The 69 DEPs were involved not only in the cell cycle, but also in cell metabolism (Figure 3A, B). Since metabolic reprogramming was a well-established hallmark of cancer, alterations in metabolism-related proteins expression were common in tumors.²⁴ According to the Figure 4, metabolically related DEPs were roughly divided into two groups: carbohydrate metabolism-related proteins and amino acid metabolism-related proteins. ENO3, FBP1, FBP2, GPD1, and ALDOB were all glycolytic pathway related proteins with inhibitory effects on tumor.²⁵⁻²⁸ For instance, ALDOB disrupted redox homeostasis by reducing the levels of fructose 1,6-bisphosphate in tumor cells, which could inhibit tumor cell proliferation.²⁶ Previous research also showed that although gluconeogenesis was frequently suppressed in tumors, re-expression of gluconeogenesis enzymes such as FBP1 could inhibit tumor growth.²⁸ As an enzyme responsible for the biosynthesis of

arginine in most body tissues, ASS1 was downregulated in multiple diverse cancers to reprogram arginine metabolism to make tumor cells more aggressive.²⁹ What's more, according to our results, these metabolism-related proteins that inhibit cancer were also down-regulated. But also as a protein related to amino acid metabolism, PYCR1 was highly expressed to maintain the redox balance of tumor cells and prevent apoptosis by synthesizing proline.³⁰ Despite the DEPs associated with metabolism and cell proliferation, quite a few DEPs were associated with the extracellular matrix. As a large extracellular matrix proteoglycan, VCAN regulated proliferation, invasion, and metastasis adhesion in a vast majority of tumor cells and VCAN expression was associated with poor prognosis in most cancers.³¹⁻³³ THBS2 was also an extracellular matrix protein and promoted cell migration and angiogenesis.³⁴ Distinguished with VCAN and THBS2, though DCN was associated with the extracellular matrix, it could antagonize many tyrosine kinase receptors to inhibit tumor development and progression.³⁵ According to these results, the four DEPs interaction clusters manifested that one cluster was involved in cell growth and division, one in carbohydrate metabolism, one in amino acid metabolism, and the rest in the extracellular matrix regulation. To summarize, the functions of the 69 DEPs fell into three main categories: cell proliferation and division, cellular metabolism, and extracellular matrix regulation.

In the following step, we performed Kaplan-Meier survival analyses of 69 DEPs one by one and found that only 10 DEPs were significantly correlated with survival for multiple cancer. Of the ten proteins, preceding text showed that some studies identified RRM2, PLOD2, MKI67, MCM5, and CKD1 promoted cancer progression and FBP1, FBP2, ENO3, GPD1, and ASS1 inhibited cancer progression, which was consistent with our results (Supplementary Figure 1). Nevertheless, this traditional way of concentrating on molecular biomarkers such as single protein has not been successful; because the development and progression of cancers were primarily accomplished by a set of biomolecules, rather than the dysfunction of an individual molecule.^{36,37} As shown in Supplementary Figure 2, the accuracy of the ten DEPs in predicting the prognosis of cancers was not high. Therefore, according to the LASSO regression method, we determined 24 DEPs: MKI67, LOXL2, PLIN4, IL33, MDK, P4HA2, AKR7A3, PLCXD3, CDK1, SRPX, PRPH, PRKAR2B, P4HA1, CALML3, SFN, DES, PHYHD1, GPD1, AADAT, PGM5, ADH1C, FBP2, ENO3, EHD3. In accordance with the above classification, among the 24 proteins, CDK1, SFN, PRKAR2B, MKI67 and MDK were involved in the cell cycle;^{16,23,38} AKR7A3, GPD1, ENO3, FBP2, AADAT, PGM5 and ADH1C were involved in cell metabolism;^{25,27,39} LOXL2, P4HA1, P4HA2, SPRX, DES, PRPH and CALML3 were involved in construction and regulation of extracellular matrix.⁴⁰⁻⁴² And most of these proteins have been identified to contribute to prognosis of many cancers.^{18,25,40-43} Although IL33 and EHD3 did not belong to any of the three groups mentioned above, some researches showed that they could inhibit the proliferation of tumor cells.^{44,45} In addition to these widely studied proteins, there were still several proteins whose roles in cancer were unclear such as PLCX3, PHYHD1 and PLIN3, which provided a new direction for cancer research. Although no research had yet explored the specific ways in which they interacted, according to correlation analysis, PGM5 was related to IL33 and DES. Therefore, we inferred that PGM5 may be involved in the regulation of tumor inflammation and extracellular matrix by regulating metabolism. Based on the 24 DEPs-based classification, we divided the cancer patients into

two groups in training cohort. The Kaplan-Meier survival analysis and the ROC analysis showed that the 24-DEPs survival-predictor model was better predictor than single protein (Figure 6B, C). We further verified the correctness of this grouping method in validation cohort and the two groups also showed significantly different survival rates (Figure 6E). Therefore, the DEPs-based survival-predictor model showed excellent survival prediction effect and is applicable to most cancers, which will contribute to therapeutic decision-making.

Yet, there are several limitations in this study. Firstly, this study mainly explored the effect of the differentially expressed proteins on predicting the OS of multiple cancers. It will inevitably be interesting to combine proteomics with genomics and even metabonomics to predict pan-cancer OS in the future. Secondly, the current study was a retrospective study utilizing the CPTAC database. Therefore, more prospective studies were still needed. Moreover, proteins data of this study were based on clinical specimens, which had limitations for clinical application. It would be clinically valuable, if we could discover tumor biomarkers in various accessible blood samples.

Conclusions

In summary, our study screened 69 differentially expressed proteins in five cancers. Then we confirmed these DEPs were mainly associated with cell proliferation and division, cellular metabolism and extracellular matrix. According to the LASSO regression method, we have determined 24 DEPs. Notably, the DEPs-based survival-predictor model could accurately predict the OS in multiple cancers. And this is the first study to utilize proteomics to construct a pan-cancer prognosis model, and the results indicated that the pan-cancer analysis may complement single cancer analysis in the identification of prognostically differentially expressed proteins.

List Of Abbreviations

CPTAC, Clinical Proteomic Tumor Analysis Consortium; HCC, hepatocellular carcinoma; LUAD, lung adenocarcinoma; CBTTC, children's brain tumor tissue consortium; CCRC, clear cell renal cell carcinoma; UCEC, uterine corpus endometrial carcinoma; DEPs, differentially expressed proteins; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; tdROC, time dependent receiver operating characteristic; AUC, area under receiver operating characteristic curve; OS, over survival

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials:

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests:

The authors declare that they have no competing interests

Funding:

This work was not supported by any funding.

Authors' contributions:

HQS designed the current study. WGH, WQW and YKX analyzed the data and wrote the manuscript. All authors read and approved the final version of the manuscript and agreed to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgments:

We are grateful to all of the reviewers for their comments

References

1. Collaborators GBDCoD. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018; 392: 1736-88.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018; 68: 394-424.
3. Cortes J, Perez-Garcia JM, Llombart-Cussac A, Curigliano G, El Saghir NS, Cardoso F, Barrios CH, Wagle S, Roman J, Harbeck N, Eniu A, Kaufman PA, Tabernero J, et al. Enhancing global access to cancer medicines. *CA Cancer J Clin*. 2020; 70: 105-24.
4. Zheng H, Zhang G, Zhang L, Wang Q, Li H, Han Y, Xie L, Yan Z, Li Y, An Y, Dong H, Zhu W, Guo X. Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis. *Front Oncol*. 2020; 10: 68.
5. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013; 41: D991-5.

6. Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015; 19: A68-77.
7. Quesada-Calvo F, Massot C, Bertrand V, Longuespee R, Bletard N, Somja J, Mazzucchelli G, Smargiasso N, Baiwir D, De Pauw-Gillet MC, Delvenne P, Malaise M, Coimbra Marques C, et al. OLFM4, KNG1 and Sec24C identified by proteomics and immunohistochemistry as potential markers of early colorectal cancer stages. *Clin Proteomics*. 2017; 14: 9.
8. Burns J, Wilding CP, R LJ, P HH. Proteomic research in sarcomas - current status and future opportunities. *Semin Cancer Biol*. 2020; 61: 56-70.
9. Noujaim J, Payne LS, Judson I, Jones RL, Huang PH. Phosphoproteomics in translational research: a sarcoma perspective. *Ann Oncol*. 2016; 27: 787-94.
10. Fumet JD, Truntzer C, Yarchoan M, Ghiringhelli F. Tumour mutational burden as a biomarker for immunotherapy: Current data and emerging concepts. *Eur J Cancer*. 2020; 131: 40-50.
11. Zhang X, Wang L, Qu Y. Targeting the beta-catenin signaling for cancer therapy. *Pharmacol Res*. 2020: 104794.
12. Zhao H, Mu X, Zhang X, You Q. Lung Cancer Inhibition by Betulinic Acid Nanoparticles via Adenosine 5'-Triphosphate (ATP)-Binding Cassette Transporter G1 Gene Downregulation. *Med Sci Monit*. 2020; 26: e922092.
13. Wu P, Heins ZJ, Muller JT, Katsnelson L, de Bruijn I, Abeshouse AA, Schultz N, Fenyo D, Gao J. Integration and Analysis of CPTAC Proteomics Data in the Context of Cancer Genomics in the cBioPortal. *Mol Cell Proteomics*. 2019; 18: 1893-8.
14. Kim SM, Kim Y, Jeong K, Jeong H, Kim J. Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography*. 2018; 37: 36-42.
15. Bradshaw RA, Hondermarck H, Rodriguez H. Cancer Proteomics and the Elusive Diagnostic Biomarkers. *Proteomics*. 2019; 19: e1800445.
16. Santamaria D, Barriere C, Cerqueira A, Hunt S, Tardy C, Newton K, Caceres JF, Dubus P, Malumbres M, Barbacid M. Cdk1 is sufficient to drive the mammalian cell cycle. *Nature*. 2007; 448: 811-5.
17. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer*. 2009; 9: 153-66.
18. Ito Y, Takeda T, Sakon M, Monden M, Tsujimoto M, Matsuura N. Expression and prognostic role of cyclin-dependent kinase 1 (cdc2) in hepatocellular carcinoma. *Oncology*. 2000; 59: 68-74.
19. Hongo F, Takaha N, Oishi M, Ueda T, Nakamura T, Naitoh Y, Naya Y, Kamoi K, Okihara K, Matsushima T, Nakayama S, Ishihara H, Sakai T, et al. CDK1 and CDK2 activity is a strong predictor of renal cell carcinoma recurrence. *Urol Oncol*. 2014; 32: 1240-6.
20. Braun KA, Breeden LL. Nascent transcription of MCM2-7 is important for nuclear localization of the minichromosome maintenance complex in G1. *Mol Biol Cell*. 2007; 18: 1447-56.

21. Chen G, Luo Y, Warncke K, Sun Y, Yu DS, Fu H, Behera M, Ramalingam SS, Doetsch PW, Duong DM, Lammers M, Curran WJ, Deng X. Acetylation regulates ribonucleotide reductase activity and cancer cell growth. *Nat Commun.* 2019; 10: 3213.
22. Nesterova M, Bossis I, Wen F, Horvath A, Matyakhina L, Stratakis CA. An immortalized human cell line bearing a PRKAR1A-inactivating mutation: effects of overexpression of the wild-type Allele and other protein kinase A subunits. *J Clin Endocrinol Metab.* 2008; 93: 565-71.
23. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, Ellis M, Henry NL, Hugh JC, Lively T, McShane L, Paik S, Penault-Llorca F, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst.* 2011; 103: 1656-64.
24. Veeranki OL, Tong Z, Mejia A, Verma A, Katkhuda R, Bassett R, Kim TB, Wang J, Lang W, Mino B, Solis L, Kingsley C, Norton W, et al. A novel patient-derived orthotopic xenograft model of esophageal adenocarcinoma provides a platform for translational discoveries. *Dis Model Mech.* 2019; 12.
25. Asberg C, Hjalmarson O, Alm J, Martinsson T, Waldenstrom J, Hellerud C. Fructose 1,6-bisphosphatase deficiency: enzyme and mutation analysis performed on calcitriol-stimulated monocytes with a note on long-term prognosis. *J Inherit Metab Dis.* 2010; 33 Suppl 3: S113-21.
26. Wang J, Wu Q, Qiu J. Accumulation of fructose 1,6-bisphosphate protects clear cell renal cell carcinoma from oxidative stress. *Lab Invest.* 2019; 99: 898-908.
27. Xie J, Ye J, Cai Z, Luo Y, Zhu X, Deng Y, Feng Y, Liang Y, Liu R, Han Z, Liang Y, Zheng Y, Mo R, et al. GPD1 enhances the anti-cancer effects of metformin by synergistically increasing total cellular glycerol-3-phosphate. *Cancer Res.* 2020.
28. Leithner K. Epigenetic Marks Repressing Gluconeogenesis in Liver and Kidney Cancer. *Cancer Res.* 2020; 80: 657-8.
29. Phillips MM, Sheaff MT, Szlosarek PW. Targeting arginine-dependent cancers with arginine-degrading enzymes: opportunities and challenges. *Cancer Res Treat.* 2013; 45: 251-62.
30. Milne K, Sun J, Zaal EA, Mowat J, Celie PHN, Fish A, Berkers CR, Forlani G, Loayza-Puch F, Jamieson C, Agami R. A fragment-like approach to PYCR1 inhibition. *Bioorg Med Chem Lett.* 2019; 29: 2626-31.
31. Liu X, Han C, Liao X, Yu L, Zhu G, Su H, Qin W, Lu S, Ye X, Peng T. Genetic variants in the exon region of versican predict survival of patients with resected early-stage hepatitis B virus-associated hepatocellular carcinoma. *Cancer Manag Res.* 2018; 10: 1027-36.
32. Salem M, O'Brien JA, Bernaudo S, Shower H, Ye G, Brkic J, Amleh A, Vanderhyden BC, Refky B, Yang BB, Krylov SN, Peng C. miR-590-3p Promotes Ovarian Cancer Growth and Metastasis via a Novel FOXA2-Versican Pathway. *Cancer Res.* 2018; 78: 4175-90.
33. Long X, Deng Z, Li G, Wang Z. Identification of critical genes to predict recurrence and death in colon cancer: integrating gene expression and bioinformatics analysis. *Cancer Cell Int.* 2018; 18: 139.
34. Bornstein P, Armstrong LC, Hankenson KD, Kyriakides TR, Yang Z. Thrombospondin 2, a matricellular protein with diverse functions. *Matrix Biol.* 2000; 19: 557-68.
35. Goldoni S, Humphries A, Nystrom A, Sattar S, Owens RT, McQuillan DJ, Ireton K, Iozzo RV. Decorin is a novel antagonistic ligand of the Met receptor. *J Cell Biol.* 2009; 185: 743-54.

36. Wang X. Role of clinical bioinformatics in the development of network-based Biomarkers. *J Clin Bioinforma.* 2011; 1: 28.
37. Gov E, Arga KY. Differential co-expression analysis reveals a novel prognostic gene module in ovarian cancer. *Sci Rep.* 2017; 7: 4996.
38. Dos Santos P, Machado ART, De Grandis RA, Ribeiro DL, Tuttis K, Morselli M, Aissa AF, Pellegrini M, Antunes LMG. Transcriptome and DNA methylation changes modulated by sulforaphane induce cell cycle arrest, apoptosis, DNA damage, and suppression of proliferation in human liver cancer cells. *Food Chem Toxicol.* 2020; 136: 111047.
39. Sun Y, Long H, Sun L, Sun X, Pang L, Chen J, Yi Q, Liang T, Shen Y. PGM5 is a promising biomarker and may predict the prognosis of colorectal cancer patients. *Cancer Cell Int.* 2019; 19: 253.
40. Liu X, Liu T, Hu L, Jiang T, Liu H, Wang Y, Lei Y, Zhu J, Bu Y. Identification and characterization of the promoter of cancer-related gene LOXL2. *Exp Cell Res.* 2020; 387: 111786.
41. Liu CN, Zhang HY, Liu CL, Wang CC. Upregulation of lncRNA CALML3-AS1 promotes cell proliferation and metastasis in cervical cancer via activation of the Wnt/beta-catenin pathway. *Eur Rev Med Pharmacol Sci.* 2019; 23: 5611-20.
42. Agarwal S, Behring M, Kim HG, Bajpai P, Chakravarthi B, Gupta N, Elkholy A, Al Diffalha S, Varambally S, Manne U. Targeting P4HA1 with a Small Molecule Inhibitor in a Colorectal Cancer PDX Model. *Transl Oncol.* 2020; 13: 100754.
43. Filippou PS, Karagiannis GS, Constantinidou A. Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene.* 2020; 39: 2040-54.
44. Xu L, Zheng Y, Wang J, Xu Y, Xie Y, Yang ZP. IL33 activates CD8+T and NK cells through MyD88 pathway to suppress the lung cancer cell growth in mice. *Biotechnol Lett.* 2020.
45. Chukkapalli S, Amessou M, Dekhil H, Dilly AK, Liu Q, Bandyopadhyay S, Thomas RD, Bejna A, Batist G, Kandouz M. Ehd3, a regulator of vesicular trafficking, is silenced in gliomas and functions as a tumor suppressor by controlling cell cycle arrest and apoptosis. *Carcinogenesis.* 2014; 35: 877-85.

Tables

Table 1. The detailed information of differentially expressed proteins for constructing the prognostic signature

Protein name	Gene name	β
alpha-aminoadipate aminotransferase	AADAT	-0.051658113
Alcohol dehydrogenase 1C	ADH1C	-0.165116243
Aflatoxin B1 aldehyde reductase member 3	AKR7A3	0.13190529
Calmodulin-like protein 3	CALML3	0.046728326
Cyclin-dependent kinase 1	CDK1	0.120550067
Desmin	DES	0.002519638
EH domain-containing protein 3	EHD3	-0.388025694
Beta-enolase	ENO3	-0.26539137
Fructose-1,6-bisphosphatase isozyme 2	FBP2	-0.245540744
Glycerol-3-phosphate dehydrogenase [NAD(+)]	GPD1	-0.041717955
Interleukin-33	IL33	0.1638571
Lysyl oxidase homolog 2	LOXL2	0.259559228
Midkine	MDK	0.153385186
Proliferation marker protein Ki-67	MKI67	0.30323553
Prolyl 4-hydroxylase subunit alpha-1	P4HA1	0.046806691
Prolyl 4-hydroxylase subunit alpha-2	P4HA2	0.144674736
Phosphoglucomutase-like protein 5	PGM5	-0.07403554
Phytanoyl-CoA dioxygenase domain-containing protein 1	PHYHD1	-0.02187378
PI-PLC X domain-containing protein 3	PLCXD3	0.121054348
Perilipin-4	PLIN4	0.216349151
cAMP-dependent protein kinase type II-beta regulatory subunit	PRKAR2B	0.063303645
Peripherin	PRPH	0.069267063
14-3-3 protein sigma	SFN	0.034228846
Sushi repeat-containing protein SRPX	SRPX	0.077423785

. Univariate and multivariate COX analyses of the DEPs-based classifier for OS.

Prognostic parameter	Univariate analysis			Multivariate analysis		
	HR	95% CI	P value	HR	95% CI	P value
CCRC						
(> 65 vs. ≤ 65)	0.724	0.218-2.409	0.599			
Sex (male vs. female)	1.099	0.297-4.063	0.887			
Grade (G3&4 vs. G1&2)	1.901	0.603-5.993	0.273			
Tumor stage (+ IV vs. I + II)	2.031	1.224-3.369	0.006			
Classification (- T4 vs. T1 + T2)	5.479	1.483-20.240	0.011			
Classification (N1 vs. N0)	1.269	0.132-12.221	0.837			
Classification (M1 vs. M0)	10.190	2.625-39.560	0.001	6.593	1.670-26.026	0.007
DEP-based classifier (High vs. Low risk)	4.047	2.081-7.871	<0.001	4.438	1.704-11.561	0.002
HCC						
(> 65 vs. ≤ 65)	0.589	0.233-1.488	0.263			
Sex (male vs. female)	0.843	0.420-1.692	0.630			
Number of Tumors (Multiple VS Single)	1.042	0.467-2.323	0.920			
Tumor thrombus (Present vs. absent)	2.118	1.157-3.879	0.015	0.780	0.399-1.525	0.468
DEP-based classifier (High vs. Low risk)	3.892	2.563-5.909	<0.001	3.114	1.911-5.073	<0.001
LUAD						
(> 65 vs. ≤ 65)	6.184	0.803-47.630	0.080			
Sex (male vs. female)	1.176	0.394-3.509	0.771			
Grade (G3&4 vs. G1&2)	0.259	0.026-2.540	0.246			
Tumor stage (+ IV vs. I + II)	7.518	2.428-23.283	<0.001	2.721	0.347-21.351	0.341
Classification (- T4 vs. T1 + T2)	4.499	1.357-14.913	0.014	2.571	0.730-9.057	0.142
Classification (N1 vs. N0)	5.400	1.658-17.585	0.005	1.973	0.257-15.160	0.514
Classification (M1 vs. M0)	0.048	0.00-10988250	0.757	6.593	1.670-26.026	0.007
Smoking history (Present vs. Absent)	2.310	0.754-7.070	0.143			
DEP-based classifier (High vs. Low risk)	3.867	1.573-9.502	0.003	2.666	1.123-6.331	0.026
UCEC						
(> 65 vs. ≤ 65)	131.682	0.007-2628475	0.334			
Grade (G3&4 vs. G1&2)	3.042	0.190-48.629	0.432			
Tumor stage (+ IV vs. I + II)	5355.564	0-5.000E+18	0.626			
Classification (- T4 vs. T1 + T2)	18.422	1.654-205.237	0.018	25773.5	0-4.44094E+23	0.424
Classification (N1 vs. N0)	0.039	0-1243745686	0.793	0.000	0.000-1102092.5	0.424
Classification (M1 vs. M0)	35.500	2.220-567.557	0.012			
DEP-based classifier (High vs. Low risk)	13.430	1.983-183.437	0.050	216.159	0.003-14902987	0.344
CBTTC						
(> 65 vs. ≤ 65)	0.945	0.895-0.998	0.044	0.971	0.926-1.019	0.228
Sex (male vs. female)	0.708	0.387-1.293	0.261			
Surgery (Present vs. Absent)	0.102	0.047-0.222	<0.001	0.165	0.070-0.387	<0.001
DEP-based classifier (High vs. Low risk)	13.430	1.983-183.437	0.050	2.173	1.229-3.843	0.008

HR, hazard ratio; CI, confidence interval; DEPs, differentially expressed proteins; hepatocellular carcinoma, HCC; children's brain tumor tissue consortium, CBTTC; clear cell renal cell carcinoma, CCRC; lung adenocarcinoma, LUAD; uterine corpus endometrial carcinoma, UCEC.

Table 3. Correlations between risk score of the DEPs-based classifier with overall survival and clinicopathological characteristics in five types of cancers.

Clinicopathological variables	Number of patients	High Risk	Low Risk	P value
LC				
Age	<65 (n, %)	67 (59.3%)	34 (30.1%)	0.323
	≥65 (n, %)	46 (40.7%)	19 (16.8%)	
Sex	Male (n, %)	30 (26.5%)	11 (9.7%)	0.190
	Female (n, %)	83 (73.5%)	42 (37.2%)	
Histologic Grade	G1+G2 (n, %)	69 (61.1%)	27 (23.9%)	0.038
	G3+G4 (n, %)	44 (38.9%)	18 (15.9%)	
	NA	0		
AJCC staging system	T1+T2 (n, %)	72 (63.7%)	29 (25.7%)	0.061
	T3+T4 (n, %)	41 (36.3%)	24 (21.2%)	
	NA	0		
	N0 (n, %)	14 (77.8%)	9 (50.0%)	
	N1 (n, %)	4 (22.2%)	3(16.7%)	
NA	95			0.688
CC				
Age	<65 (n, %)	120 (85.1%)	85 (60.3%)	0.615
	≥65 (n, %)	21 (14.9%)	16 (11.3%)	
Sex	Male (n, %)	26 (18.4%)	19 (13.5%)	0.856
	Female (n, %)	115 (81.6%)	82 (58.2%)	
Number of Tumors	Single (n, %)	121(85.8%)	87 (61.7%)	0.861
	Couple (n, %)	20 (14.2%)	14 (9.9%)	
	NA	0		
AD				
Age	<65 (n, %)	59 (57.8%)	29 (28.4%)	0.466
	≥65 (n, %)	43 (42.2%)	18 (17.6%)	
Sex	Male (n, %)	32 (31.4%)	12 (11.8%)	0.240
	Female (n, %)	70 (68.6%)	35 (34.3%)	
Histologic Grade	G1+G2 (n, %)	62 (63.9%)	26 (26.8%)	0.367
	G3+G4 (n, %)	35 (36.1%)	18 (18.6%)	
	NA	5		
AJCC staging system	T1+T2 (n, %)	91 (89.2%)	39 (38.2%)	0.060
	T3+T4 (n, %)	11 (10.8%)	8 (7.8%)	
	NA	0		
	N0 (n, %)	70 (68.6%)	29 (28.4%)	
	N1 (n, %)	32 (31.4%)	18 (17.6%)	
NA	0			0.163
II (n, %)	85 (97.7%)	42 (48.3%)	43 (49.4%)	0.167
III (n, %)	2 (2.3%)	0 (.0%)	2 (2.3%)	
NA	25			
Histological stage	II (n, %)	81 (79.4%)	34 (33.3%)	0.103
	III (n, %)	21 (20.6%)	13 (12.7%)	
	NA	0		
Smoking history	Present (n, %)	56 (56.6%)	26 (26.3%)	0.993
	Absent (n, %)	43 (43.4%)	20 (20.2%)	
NA	3			
EC				
Age	<65 (n, %)	56 (56.6%)	16 (16.2%)	0.251
	≥65 (n, %)	43 (43.4%)	8 (8.1%)	
Histologic Grade	G1+G2 (n, %)	73 (75.3%)	13 (13.4%)	0.006
	G3+G4 (n, %)	24 (24.7%)	11 (11.3%)	

NA	2				
staging system					
T1+T2 (n, %)	88 (88.9%)		20 (20.2%)	68 (68.7%)	
T3+T4 (n, %)	11 (11.1%)		4 (4.0%)	7 (7.1%)	0.320
NA	0				
N0 (n, %)	47 (85.5%)		9 (16.4%)	38 (69.1%)	
N1 (n, %)	8 (14.5%)		4 (7.3%)	4 (7.3%)	0.058
NA	44				
0 (n, %)	71 (97.3%)	17 (23.3%)		54 (74.0%)	
1 (n, %)	2 (2.7%)	1 (1.44%)		1 (1.4%)	0.399
NA	26				
stage					
II (n, %)	82 (82.8%)	18 (18.2%)		64 (64.6%)	
IV (n, %)	17 (17.2%)	6 (6.1%)		11 (11.1%)	0.243
NA	0				
TC					
<65 (n, %)	195 (99.5%)		100 (51.0%)	95 (48.5%)	
≥65 (n, %)	1 (0.5%)		1 (0.5%)	0 (0%)	0.331
der					
Male (n, %)	86 (43.9%)		44 (22.4%)	42 (21.4%)	
Female (n, %)	110 (56.1%)		57 (29.1%)	53 (27.0%)	0.927
ger					
Present (n, %)	163 (93.1%)		82 (46.7%)	81 (46.3%)	
Absent (n, %)	12 (6.9%)		10 (5.7%)	2 (1.1%)	0.061
N	21				

Hepatocellular carcinoma, HCC; children's brain tumor tissue consortium, CBTTC; clear cell renal cell carcinoma, CCRC; lung adenocarcinoma, LUAD; uterine corpus endometrial carcinoma, UCEC.

Supplementary Figure Legends

Supplementary Figure 1: Kaplan–Meier analysis of 10 DEPs in the training cohort. (A) RRM2; (B) PLOD2; (C) MKI67; (D) MCM5; (E) CDK1; (F) FBP1; (G) FBP2; (H) ENO3; (I) GPD1; (J) ASS1. All log rank P values of 10 DEPs were lower than 0.05. DEPs, differentially expressed proteins.

Supplementary Figure 2: ROC analysis of 10 DEPs in the training cohort. (A) RRM2; (B) PLOD2; (C) MKI67; (D) MCM5; (E) CDK1; (F) FBP1; (G) FBP2; (H) ENO3; (I) GPD1; (J) ASS1. DEPs, differentially expressed proteins; AUC, Area Under the Curve.

Figures

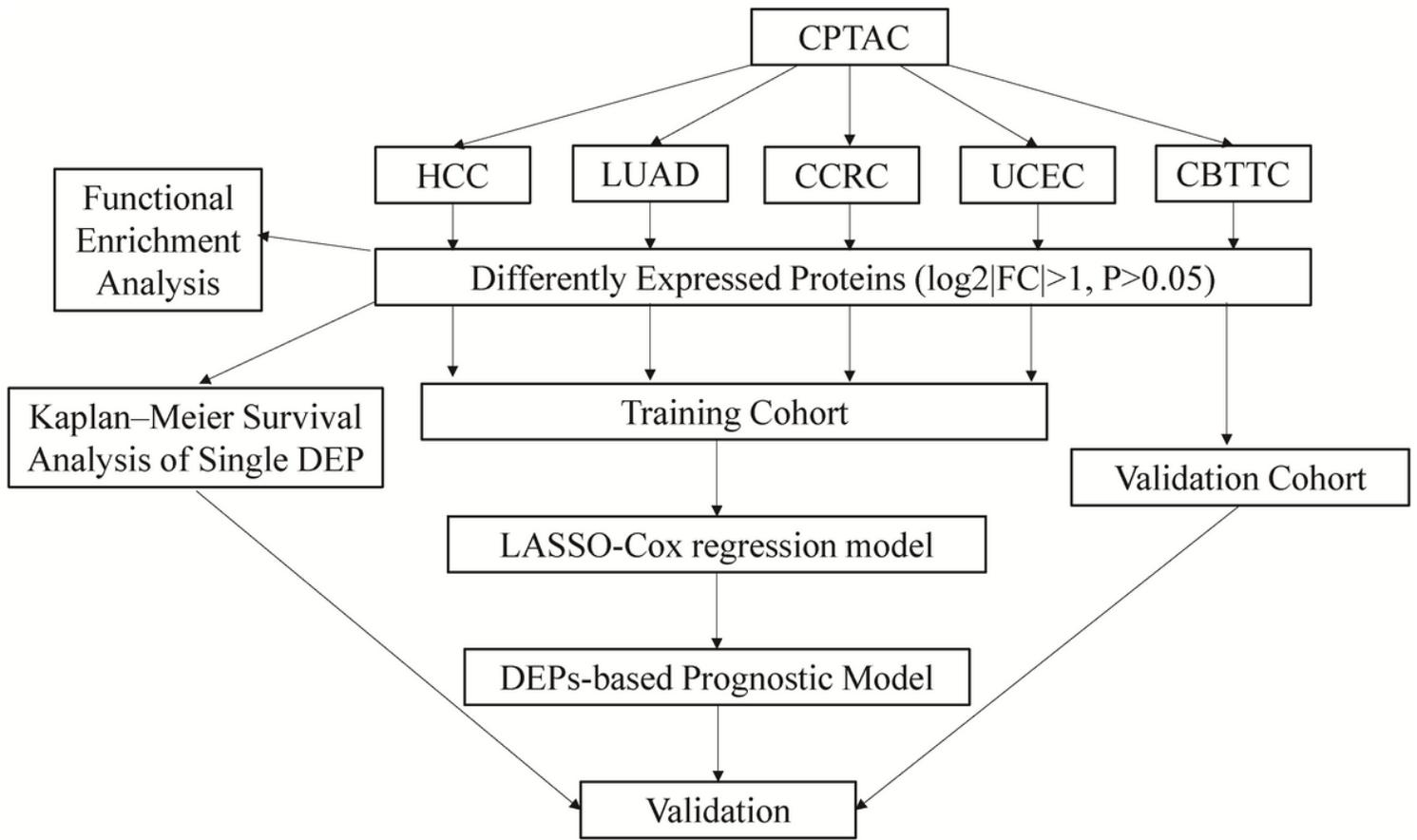


Figure 1

The workflow of this work.

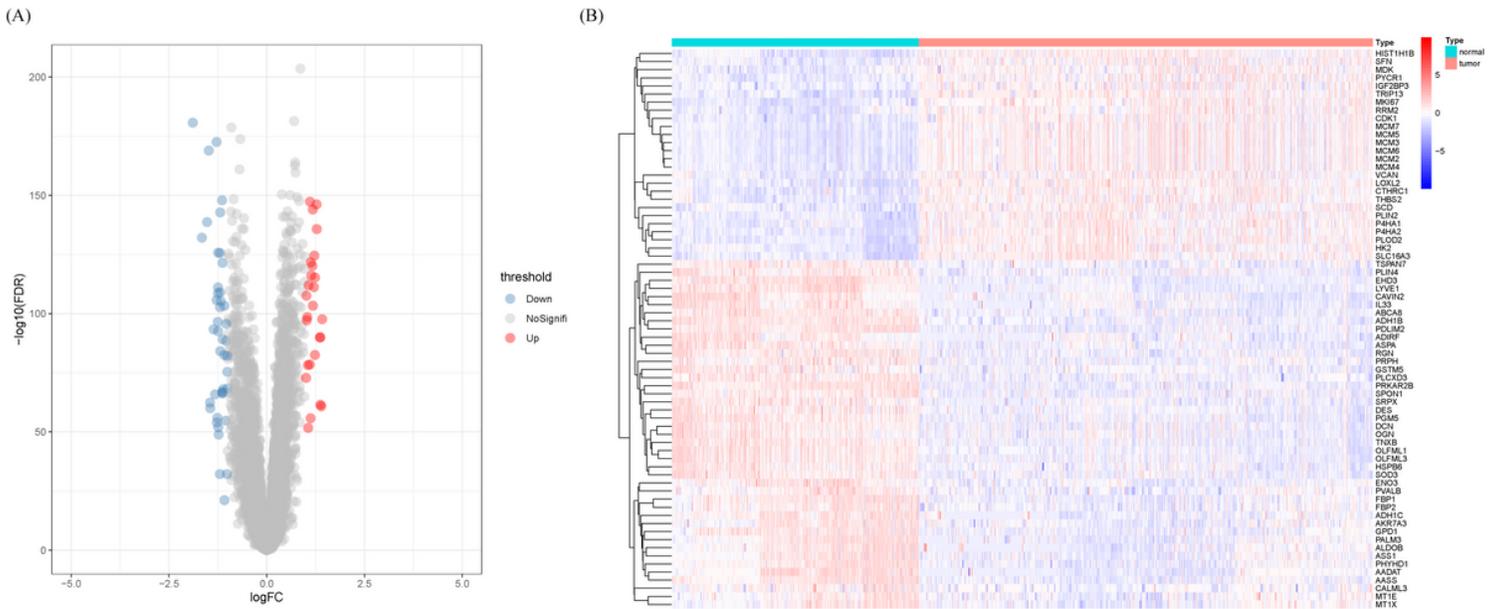


Figure 2

Identification of DEPs in five cancers. DEPs were defined with P-value < 0.05 and $|\log_2(\text{Fold Change})| > 1$. (A) Volcano plots of proteins with normalized expression alteration in all five cancers; (B) Heatmap of the

DEPs (n=69) in all five cancers. DEPs, differentially expressed proteins.

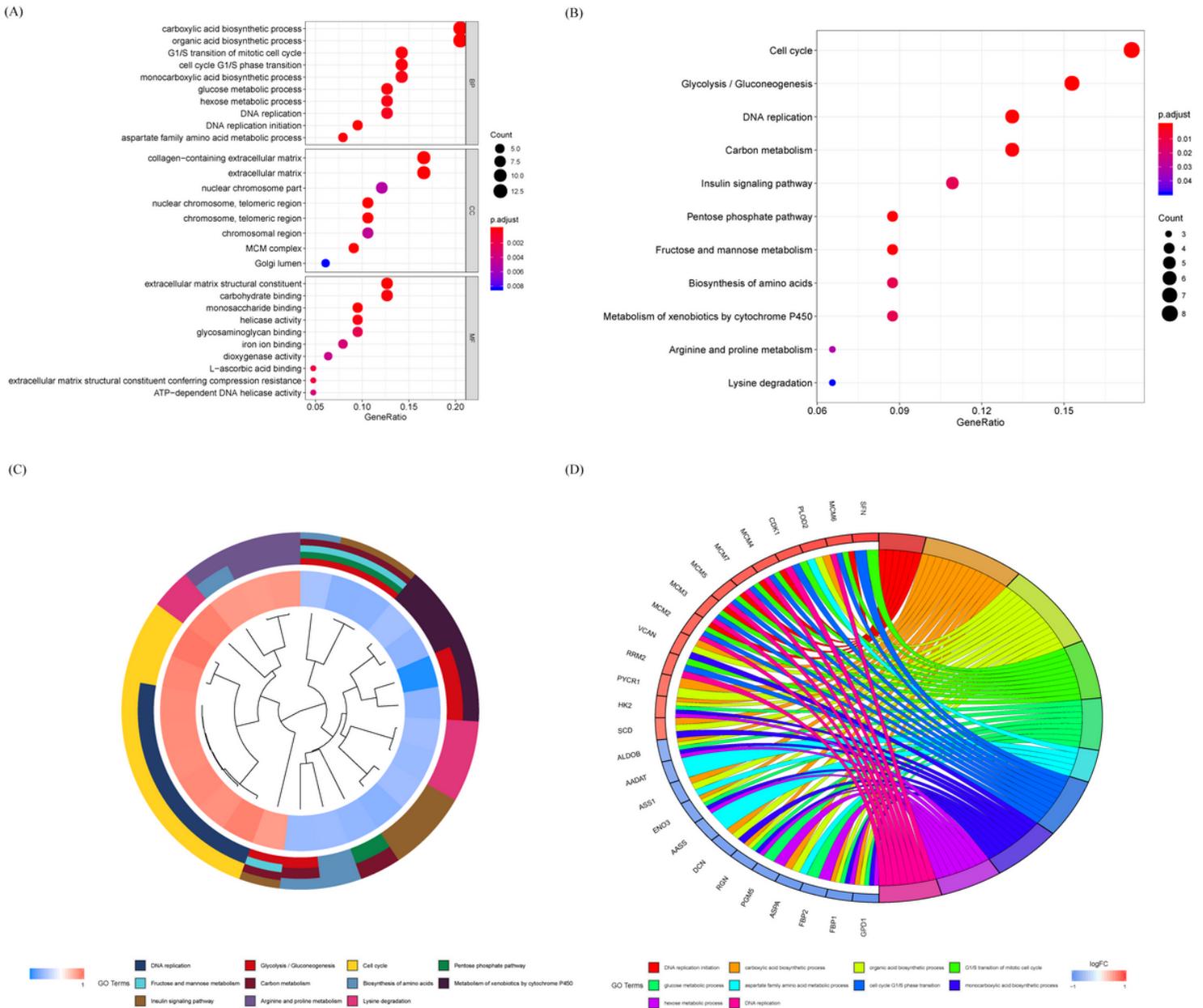


Figure 3

GO analysis and KEGG analysis of DEPs. (A) The functions of the 69 DEPs identified cover three main categories: BP, CC, MF; (B) based on KEGG pathway, 11 enriched pathways with lowest P-value were displayed; (C) (D) GO cluster diagram and GO chord diagram of the 69 DEPs. DEPs, differentially expressed proteins; BP, biological processes; CC, cellular contents; MF, molecular functions; GO, gene ontology.

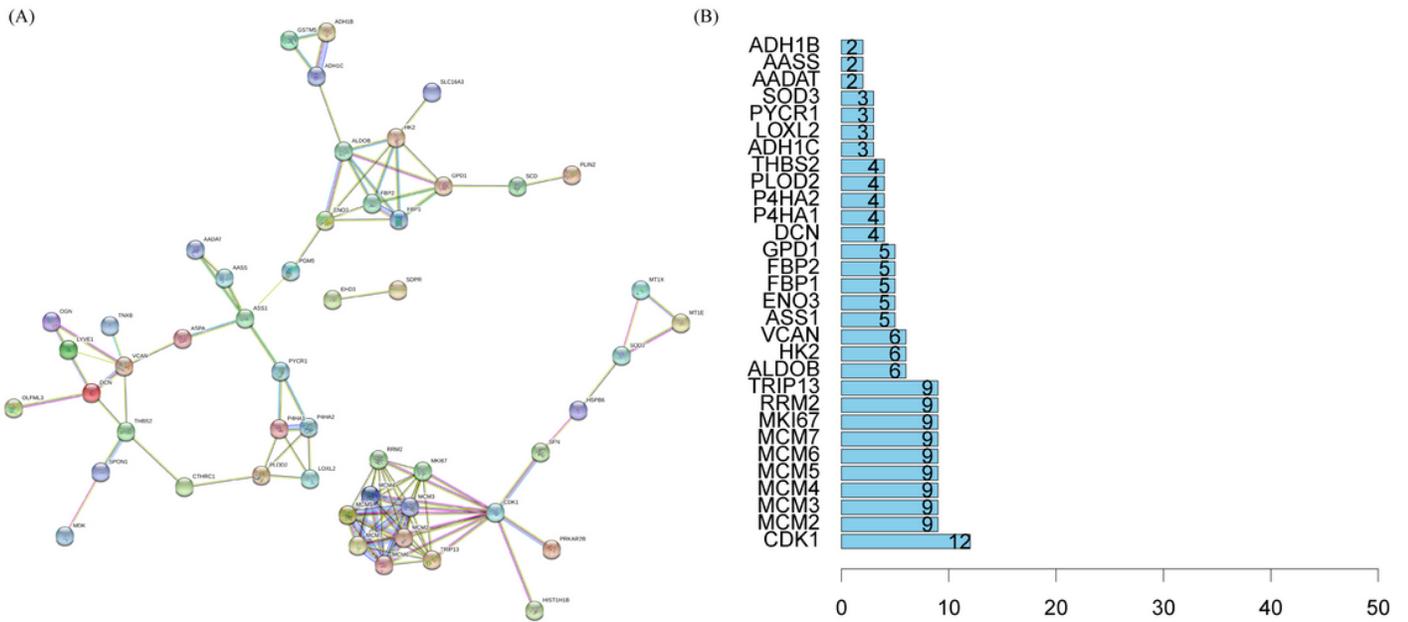


Figure 4

PPI network. (A) Interactions among 69 DEPs were detected after removing isolated proteins; (B) the number of interactions between each protein and other proteins. PPI, protein-protein interaction; DEPs, differentially expressed proteins.

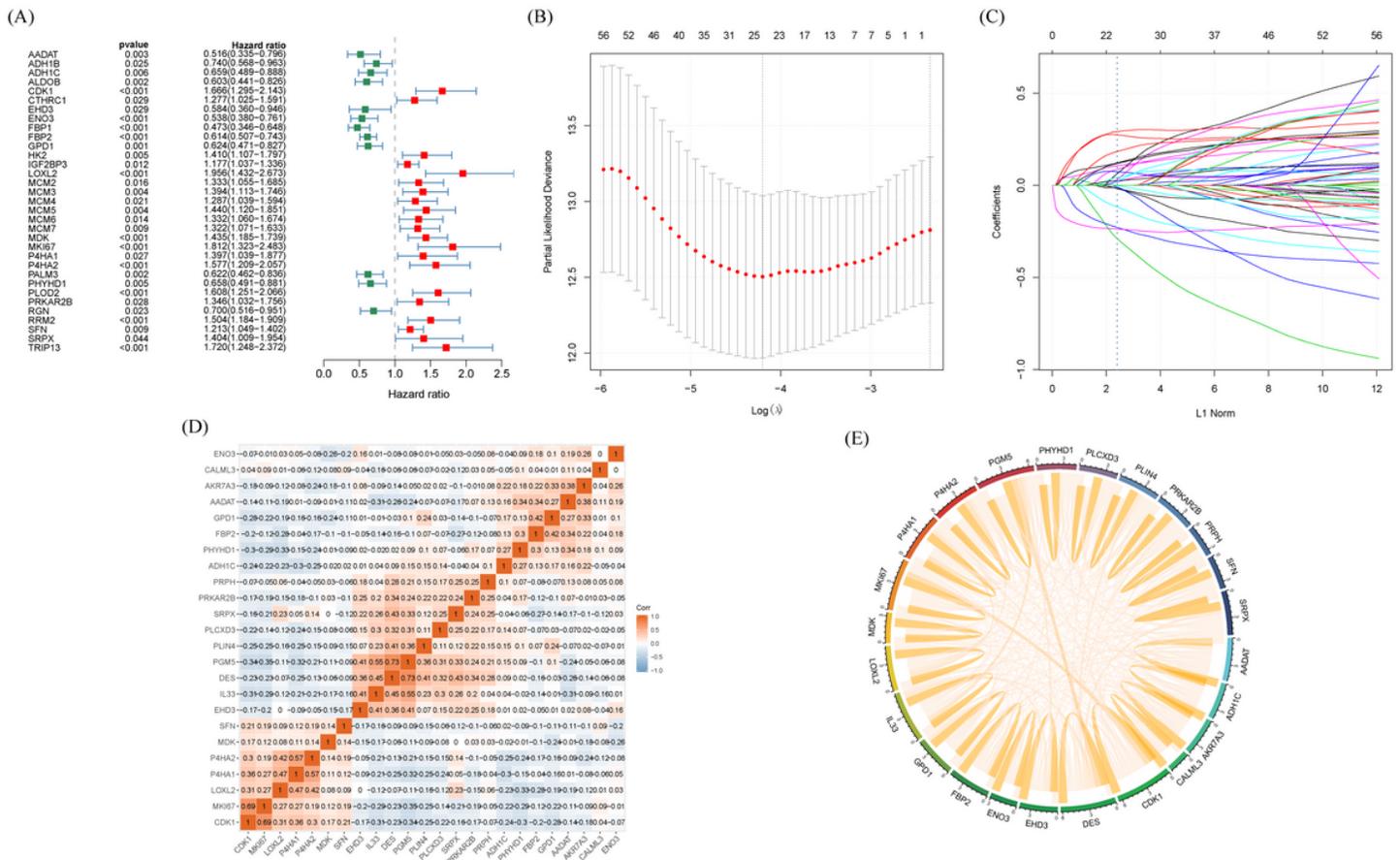


Figure 5

- [FigureS2.tif](#)
- [FigureS1.tif](#)